

Estimation of Parameters of Structural Model for Genetic Covariances in International Genetic Evaluations

R. Rekaya, K. Weigel, and D. Gianola

Department of Dairy Science

University of Wisconsin, Madison, 53706, USA

Introduction

In international genetic evaluations, genetic correlations between countries are computed using deregressed proofs as dependent variables (Sigurdsson et al., 1996). This method reduces considerably the complexity of problem from a computational and practical point of view. However, a major problem facing international genetic evaluations is the increasing number of traits (countries), and genetic ties between countries are often weak. This situation leads to problems for the computations of the genetic correlations, and imprecise correlation estimates can affect predicted genetic values. This problem becomes worse with the addition of new countries with small populations and poor genetic connections with existing INTERBULL member countries

In international genetic evaluations, traits are defined according to country borders. However similarity between herds in different countries depends not only on the geographical proximity of the pair of countries but also on genetic ties between countries, climate conditions, and management practices. Such information can be used to predict genetic covariances by relating these factors to the genetic covariance structure.

A structural model for genetic covariances that uses external data information and contain fewer parameters can lead to more precise estimates of the genetic covariance matrix between countries.

The objective of this study was to implement a structural model for genetic covariances and compare it with a standard multiple-trait model using lactation records for milk yield from nine INTERBULL member countries.

Material and Methods

Data

Data were first lactation milk yield records for daughters of AI sires in 8 Interbull member countries (Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, Israel, and Switzerland), and 5 geographical regions of the US (Midwest, Northeast, Northwest, Southeast, and Southwest). After all edits (≥ 10 records per sire, $275 \leq \text{days in milk} \leq 375$, sires with relatives in at least in two countries, and date of calving from 1979-1998). The data file consisted of 2,885,117 lactation from 17,867 sires. A summary description of the data set is reported in Table 1. The pedigree file included 25,764 animals.

Methods

Three analyses were performed in this study. The first analysis involved estimation of genetic covariances for milk yield in the thirteen countries (regions) using a standard multiple-trait approach. In the last two analyses, structural models were assumed for genetic covariances. In all analyses, the following sire model was used:

$$b_l DIM + s_{ml} + e_{ijklm}$$

where, y_{ijklm} is milk yield of a daughter of sire m in country (region) l ($l=1,13$), HY_i is the effect of the herd-year i , SC_j is the effect of calving season j ($j=1,4$) (seasons were defined by three-month periods starting in December), AC_k ($k=1,4$) is the age at calving effect, b_l is a regression coefficient, DIM is days in milk, s_{ml} is the additive genetic effect for sire m in region l and e_{ijkl} is the residual term. All systematic effects were nested within traits.

Structural models for the genetic covariances

Two models were used to describe the genetic covariances. The first (SM1) was a three-parameter model. The second (SM2) had four parameters. The 78 off-diagonal elements of the genetic covariance matrix were explained by three and four parameters, respectively in models SM1 and SM2. The covariance between two countries (regions) was written as a linear function of a set of explanatory variables

$$g_{ij} = \zeta_{ij}b$$

where **b** is a set of factors that explain the covariance between trait i and j and $\zeta_{i,j}$ is the corresponding incidence matrix.

Model SM1

The following model was used to describe the genetic covariance $g_{i,j}$ between countries (regions) i and j:

$$g_{ij} = \mu + b_1GS_{ij} + b_2MS_{ij}$$

where μ is an intercept common to all off-diagonal elements of the genetic covariance matrix, GS_{ij} and MS_{ij} are measures of genetic similarity management similarity between regions i and j, respectively, and b_1 and b_2 are regression coefficients. Following previous notation $b = (\mu, b_1, b_2)$.

Genetic similarity between regions i and j was defined as the ratio between the number of daughters of common bulls used in the two regions and the total number of daughters of all bulls. Management similarity was defined as the ratio between the absolute value of the difference between average milk yield in regions i and j and the sum of both averages.

Model SM2

Model SM2 was similar to model SM1 but an extra factor related to climate similarity was included to describe the genetic covariance g_{ij} between countries (regions) i and j

$$g_{ij} = \mu + b_1GS_{ij} + b_2MS_{ij} + b_3CS_{ij}$$

where μ , GS_{ij} and MS_{ij} are as before and CS_{ij} is a measure of climate similarity given by: $CS(i, j) = |H_i - H_j| / (H_i + H_j)$,

where H_i and H_j are heat indices for the month of July in region i and j, respectively, and $b = (\mu, b_1, b_2, b_3)$. The heat indices were calculate as $H_i = T_i - R_i/20$, where T_i and R_i were the average temperature and total rainfall in July in region i.

A Bayesian analysis via Markov chain Monte Carlo (MCMC) methods was carried out for the three models. Gibbs sampling was used for the first model (standard multiple-trait model). For models MS1 and MS2, the Metropolis and Metropolis-Hastings algorithms within the Gibbs sampling were used for the marginalization process. The Metropolis and Metropolis-Hastings algorithms were used because some conditional distributions were not closed-form (the conditional distributions of the parameters of the structural model for the covariance, b , as well as the diagonal elements of the genetic covariance matrix). For a full description of the derivation of the conditional distributions and implementation of a the structural model for genetic covariance see Rekaya et al. (1999).

For the three analyses a unique long chain of 100,000 samples was implemented. The first 20,000 samples were discarded as burn-in and the remaining samples were used in the computation of summaries from the posterior distributions of models parameters.

Results and discussion

No significant differences were observed between model SM1 and SM2, mainly because the heat index didn't help in explaining the genetic covariances. For this reason, the discussion will focus on the results of the standard multiple-trait analysis and the results of the structural model SM1.

Table 2 shows the posterior means and standard deviations of the parameters of the structural model SM1. As expected, there was

a positive relationship between the genetic covariance and genetic similarity, indicating that countries (or regions) that used the same sires tend to have higher genetic correlations. The estimate of b_2 indicated a decrease in genetic covariances when management similarity increased. However, the definition of management similarity in this study decreases when the management practices between a pair of countries (regions) are similar. This means that there is actually a positive relationship between management similarity and genetic covariances.

Table 3 shows the posterior means of the genetic variances for milk yield in the thirteen countries (regions) using the standard multiple trait model and the structural model. In general, absolute difference on the posterior means of the genetic variances between both analyses were $\leq 7\%$. The major differences were observed for Israel, Austria, and Belgium. Compared with recent estimates used by INTERBULL, the genetic variances estimated in both analyses were lower.

The residual variances (Table 4) tended to be slightly lower when the structural model for genetic covariances was used. The residual variances for milk yield in the five regions in the US were very similar to those found by Rekaya et al., 1999. As a consequence of the small changes in the genetic and residual variances for milk yield in the thirteen countries (regions), the posterior means of the heritabilities using both models were very similar. The highest heritabilities were for the European countries and Israel. Estimates of heritability of milk yield in the five regions of the US were very similar (0.27 – 0.28).

Table 8 shows the genetic correlations between the thirteen countries (regions) using the standard multiple-trait model and the structural model. With both models, the genetic correlations found in this study were lower to those published by INTERBULL. Estimates obtained using the structural model tended to be slightly lower than those found using the standard multiple-trait analysis. However some differences were large. Table 6 shows differences in the genetic correlations greater than 0.05. The major differences in

genetic correlations were between Estonia and Israel (0.11), Israel and Austria (0.08), and Austria and Estonia (0.06). All pairs of countries with major changes in genetic correlations between the two analyses shared very few sires. As a result of weak genetic ties between the pairs of countries, estimates of the genetic correlations using the standard multiple-trait were very imprecise, and point estimates may be of a little practical interest. This imprecision is shown clearly in Table 7 where the 2.5% and 97.5% quartiles of the posterior distributions of correlations are shown as a function of genetic ties (number of common sires). For the standard multiple-trait analysis, the interval from 2.5% to 97.5% is very large with a small number of common sires and it decreases as genetic ties increases. The same pattern was observed when a structural model was used. However, the 2.5% to 97.5% interval was much smaller with the structural model for the same pairs of countries indicating to a more precise estimates.

The ratio between the deviance information criterion obtained under both models was 1.07 in favor of the structural model. This result is in concordance with the pattern observed in the residual variances.

Finally, the use of structural model induced an increase in the computational cost of 15% to 20%.

Conclusions

A structural model with few parameters (three) was able to explain the genetic covariances between thirteen countries (regions), and gave a slightly better results compared with the standard multiple-trait model. The performance of the structural model could be better if more informative information on management practices and climate were available. However, In the international genetic evaluation context, it going to be difficult to increase the number of effects/variables to explain the genetic covariances, because in some countries very little management and climate information is recorded and there is a lack of harmony between countries in the recorded information.

The use of test day data in the national genetic evaluation of some INTERBULL members will increase rapidly the number of traits and the problems of inferences on the genetic covariance matrix. A structural model accounting for the temporal (in time) and spatial (location, management practices, climate, ...etc) variation can be a useful tool.

Acknowledgments

Data were generously provided by the Federation of Austrian Cattle Breeders, Vlaamse Rundveeteelt Vereniging (Belgium), Czech-Moravian Breeders Corporation, The Danish Agricultural Advisory Centre, the Finnish Animal Breeding Association, A R. O. The Volcani Center (Israel), the Holstein Association of Switzerland, the Interbull Centre (Sweden), and the USDA Animal Improvement Programs Laboratory, Beltsville,

Maryland. Financial support was provided by the National Association of Animal Breeders, Columbia, Missouri.

References

- Rekaya, R., Weigel, K. & Gianola, D. 1999. Bayesian estimation of parameters of a structural model for genetic covariances between milk yield in 5 regions of the USA. Proc. 50th Annu. Mtg. EAAP, Zurich, Switzerland, August 23-26.
- Sigurdsson, A., Banos, G. & Philipsson, J. 1996. Estimation of genetic (co)variance components for international evaluation of dairy bulls. Acta Agri. Scand., Sect. A, Animal Sci. 46, 129-136.

Table 1. Summary of the data used in the present study

Country or Region	No. Sires	No. Cows
Austria	229	25,842
Belgium	1431	158,537
Czech Republic	2105	148,390
Denmark	463	1,304,605
Estonia	248	10,429
Finland	183	27,665
Israel	107	78,970
Switzerland	231	116,614
US-Midwest	7442	260,369
US-Northeast	10043	598,970
US-Northwest	7733	234,520
US-Southeast	5032	74,918
US-Southwest	5468	76,086
Total	17,867*	3,034,933

* number of unique sires.

Table 2. Posterior means and standard deviations of the structural model (SM1) for genetic covariances

Parameter	Mean	SD
μ	51229	809
b_1	68392	1349
b_2	-7173	535

Table 3. Posterior means of genetic variances for milk yield in the thirteen countries (regions) using the standard and structural models

Country or Region	Standard	SM1
Austria	58,267	64,298
Belgium	66,928	71,202
Czech Republic	52,205	49,456
Denmark	74,757	77,383
Estonia	47,329	46,340
Finland	69,706	68,287
Israel	82,224	88,278
Switzerland	52,999	51,769
US-Midwest	89,272	88,527
US-Northeast	82189	81,642
US-Northwest	107,783	103,461
US-Southeast	83,902	84,093
US-Southwest	84279	87,194

Table 4. Posterior means of residual variances for milk yield in the thirteen countries (regions) using the standard and structural models

Country or Region	Standard	SM1
Austria	751,832	716,194
Belgium	724,251	724,921
Czech Republic	623,324	623,015
Denmark	697,332	697,712
Estonia	599,934	595,074
Finland	775,052	774,235
Israel	982,549	981,023
Switzerland	535,885	557,278
US-Midwest	1,221,086	1,221,863
US-Northeast	1,093,939	1,084,678
US-Northwest	1,311,580	1,323,587
US-Southeast	1,135,940	1,134,859
US-Southwest	1,167,647	1,165,140

Table 5. Posterior means of heritabilities for milk yield in the thirteen countries (regions) using the standard and structural models

Country or Region	Standard	SM1
Austria	0.29	0.33
Belgium	0.34	0.34
Czech Republic	0.30	0.31
Denmark	0.38	0.40
Estonia	0.28	0.27
Finland	0.33	0.33
Israel	0.31	0.34
Switzerland	0.36	0.34
US-Midwest	0.28	0.27
US-Northeast	0.28	0.28
US-Northwest	0.30	0.29
US-Southeast	0.26	0.27
US-Southwest	0.27	0.29

Table 6. Absolute differences greater than .05 in estimated genetic correlations between the two models

Country or Region	Standard	SM1	Common sires
Estonia – Israel	0.74	0.63	0
Israel – Austria	0.74	0.69	1
Austria – Estonia	0.79	0.73	3
Estonia – Finland	0.84	0.79	1
Switzerland – Israel	0.83	0.78	1
Austria – Estonia	0.77	0.72	1
Czech Republic – Finland	0.82	0.77	3
Switzerland – Estonia	0.83	0.78	1
Czech Republic - Estonia	0.75	0.70	5

Table 7. Quantiles corresponding to 2.5% and 97.5% of the posterior distribution of genetic correlations between pairs of countries with weak, average and strong genetic ties

Country or Region	Standard		SM1		Common sires
	2.5%	97.5%	2.5%	97.5%	
Estonia – Israel	0.51	0.88	0.54	0.71	0
Israel – Austria	0.53	0.85	0.57	0.75	1
Estonia – Finland	0.63	0.91	0.68	0.86	1
Denmark –	0.78	0.96	0.85	0.95	100
Northeast	0.79	0.96	0.87	0.96	337
Belgium –	0.85	0.98	0.90	0.98	1275
Northeast					
Midwest -					
Northeast					