Application of a Structural Model for Genetic Covariances in International Dairy Sire Evaluations

R. Rekaya, K. A. Weigel, and D. Gianola
Department of Dairy Science
University of Wisconsin, Madison 53706

ABSTRACT

A structural model for estimation of genetic covariances in international dairy sire evaluations was compared with a standard multiple-trait mixed model using milk yield data from 13 environments (Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, Israel, Switzerland and five regions of the United States). Data consisted of 2,885,117 first-lactation records from daughters of 17,867 Holstein sires. Parameters of the structural model for genetic covariances included an intercept and measures of genetic, management, and climate similarity. Genetic similarity was defined as the ratio between the number of daughters of common bulls used in a given pair of regions and the total number of daughters of all bulls. The measure of management similarity was the ratio between the absolute value of the difference in average milk yield between two regions and the sum of these averages. Climate similarity was a function of the difference in heat indices between pairs of regions. Results showed that the structural model for the genetic covariances gave more precise estimates of the latter than the standard multivariate analysis because the number of parameters to be estimated in the genetic covariance matrix was reduced from 91 to 16. A comparison of the two models using a deviance information criterion (a measure of quality of fit) showed a slight superiority for the structural model. For the latter, the posterior means of the genetic correlations between environments ranged between 0.63 and 0.98.

Key words: structural model, covariance components, multivariate analysis, Gibbs sampling

INTRODUCTION

In international dairy sire evaluations, genetic correlations between countries are estimated by deregressed proofs in different countries as input data in a multiple-trait analysis (Sigurdsson et al., 1996). This ad hoc method considerably reduces the complexity of the estimation problem, at least from a computational point of view because the number of progeny-tested bulls is roughly 500-fold smaller than the number of cows with records. However, major challenges facing international genetic evaluation are posed by an increasing number of countries (traits) and by the weak genetic ties or links between some of these countries. This can lead to imprecise estimates of genetic correlations between countries, which in turn, can produce EBV with low reliability. Many new International Bull Evaluation Service (INTERBULL) member countries have small populations and weak genetic connections with member countries that have historically participated in global dairy cattle breeding service.

In international dairy sire evaluations, traits are currently defined according to country borders. For example, milk yields in Belgium and in the Netherlands are treated as different traits. However, similarity in production systems between herds in different countries depends not only on geographical proximity but also on climate conditions, on management practices, and on the genetic composition of the cow population. If information is available about these variables, it can be used to explain the genetic covariance structure between countries.

A well-designed structural model for genetic covariances uses information, external to the data, to explain genetic covariability in terms of few parameters, leading to more precise estimates of genetic correlations between countries than a standard multiple-trait approach. Rekaya et al. (1999a) developed a structural model for genetic covariances and applied it to example data from five regions of the United States. Estimates of genetic variances and covariances are more precise
with the structural model. The objective of this study was to fit a structural model for genetic covariances between countries in the context of international dairy sire evaluations, and to compare it with a standard multiple-trait model. Individual lactation records for milk yield from 13 production environments in nine countries were used in the analysis, as opposed to the regressed sire proofs employed in current practice.

**MATERIALS AND METHODS**

**Data**

Data were first-lactation milk yield records of daughters of AI sires in eight INTERBULL member countries (Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, Israel, and Switzerland), and in five geographical regions of the United States (Midwest, Northeast, Northwest, Southeast, and Southwest). After all edits (≥10 records per sire, 275 ≤ DIM ≤ 375, sires with daughters in at least two countries, and calving date from 1979 to 1998), the data file contained 2,885,117 first-lactation records from progeny of 17,867 sires. A summary of the dataset is shown in Table 1.

**Methods**

Three models for estimating genetic covariances between the 13 environments were applied. The first was a standard multivariate model for estimating genetic covariances in which milk yield in each of the 13 environments was considered to be a distinct trait. The second and third models were structural models for genetic covariances, and these will be described later. In all analyses, the following sire model was used:

\[
y_{ijklmn} = \text{HY}_{i(l)} + \text{SC}_{j(l)} + \text{AC}_{k(l)} + b_1\text{DIM} + s_{m(l)} + e_{ijklmn}
\]

where: \(y_{ijklmn}\) is the milk yield of a daughter of sire \(m\) in environment \(l\) (\(l = 1,2,\ldots,13\)); \(\text{HY}_{i(l)}\) is the effect of herd-year \(i\), \(\text{SC}_{j(l)}\) is the effect of season of calving \(j\) (\(j = 1,2,\ldots,4\)) (seasons were defined by 3-mo periods starting in December); \(\text{AC}_{k(l)}\) is the effect of age at calving class \(k\) (\(k = 1,2,\ldots,4\)); \(b_1\) is a regression coefficient on DIM; \(s_{m(l)}\) is the transmitting ability of sire \(m\) in environment \(l\), and \(e_{ijklmn}\) is a residual term. All systematic effects were peculiar to production environments, as indicated in the notation.

**Structural Models for the Genetic Covariances**

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

\[
g_{ij} = k'_{ij} b
\]

where \(b\) is a vector containing the effects of factors affecting the covariance between traits \(i\) and \(j\) and \(k'_{ij}\) is the corresponding incidence row vector.

**Model SM1.** In model SM1, the genetic covariance \(g_{ij}\) between environments \(i\) and \(j\) was described as:

\[
g_{ij} = \mu + b_1\text{GS}_{ij} + b_2\text{MS}_{ij}
\]

where: \(\mu\) is an intercept common to all off-diagonal elements of the genetic covariance matrix; \(\text{GS}_{ij}\) and \(\text{MS}_{ij}\) are measures of genetic and management similarity between regions \(i\) and \(j\), respectively, and \(b_1\) and \(b_2\) are corresponding regression coefficients. Hence, \(b = (\mu, b_1, b_2)^T\).

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 yields in regions \(i\) and \(j\) and the sum of these averages. Note that \(\text{MS}_{ij}\) is a contrary measure of similarity: as \(\text{MS}_{ij}\) increases, similarity decreases.

**Model SM2.** Model SM2 was similar to SM1, but with an additional factor related to climate similarity.
included in the description of the genetic covariance $g_{ij}$ between environments $i$ and $j$. The model was:

$$g_{ij} = \mu + b_1 GS_{ij} + b_2 MS_{ij} + b_3 CS_{ij},$$  \hspace{1cm} [4]$$

where $\mu$, $GS_{ij}$ and $MS_{ij}$ are as defined previously, $CS_{ij}$ is a measure of climate similarity, and $b_1$, $b_2$ and $b_3$ are regression coefficients. In this study:

$$(CS(ij) = \frac{|H_i - H_j|}{(H_1 + H_2)},$$

where $H_i$ and $H_j$ are heat indices for the month of July in environments $i$ and $j$, respectively. The heat indices were calculated as $H_i = T_i - R_i / 20$, where $T_i$ and $R_i$ were the average temperature and total rainfall for July in environment $i$, respectively. Again, $CS(ij)$ is a contrary measure of similarity: it decreases as heat indices are more similar to each other.

**Implementation**

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). In models SM1 and SM2, the Metropolis and Metropolis-Hastings algorithms within Gibbs sampling were employed to draw samples for some parameters. We did this because in the structural model the conditional posterior distributions of parameters $b$, and of the diagonal elements of the genetic covariance matrix, are not in closed form. For a full description of the derivation of the conditional distributions and implementation of the structural model, see Rekaya et al. (1999a, 1999b).

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

**Model Comparison**

The three models were compared using the Deviance Information Criterion (DIC) as defined by Spiegelhalter et al. (1998). The criterion encompasses both measures of fit and of model complexity. It is defined as:

$$DIC = \bar{D} + p_D$$  \hspace{1cm} [5]\]

where:

$$\bar{D} = E_{\tilde{\theta}}[D(\theta)]$$  \hspace{1cm} [6]\]

is the posterior expectation of the Bayesian deviance $D(\theta) = -2 \log p(y | \theta)$  \hspace{1cm} [7]\]

and $p_D$, the “effective number of parameters”, is:

$$p_D = \bar{D} - D(\bar{\theta}),$$  \hspace{1cm} [8]\]

where $\bar{\theta}$ is the posterior mean of parameters intervening in the sampling model [1]. A smaller value of DIC indicates a better fit of the model. For a more elaborate explanation of DIC, see Spiegelhalter et al. (5).

**RESULTS AND DISCUSSION**

Table 2 gives values of similarity variables between pairs of environments. The minimum genetic similarity was between Estonia and Israel. Estonia and the northeast United States were the most dissimilar pair of environments for the management variable. Austria and the southeast United States were the most dissimilar for heat index variable.

No obvious differences were observed between models SM1 and SM2, mainly because the heat index did not help to explain the genetic covariance structure. For this reason, discussion will focus on results from the standard multiple-trait analysis and from the structural model SM1.

Table 3 shows the posterior means and standard deviations of parameters of the structural model SM1. As expected, there was a positive relationship between genetic covariance and genetic similarity, indicating that production systems using the same sires in the

---

**Table 2.** Minimum and maximum values of genetic, management, and climate similarity variables for the 13 environments.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Pair of environments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic similarity</td>
<td>0</td>
<td>Estonia–Israel</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.44</td>
<td>US–Midwest–US–Northwest</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.009</td>
<td>Israel–US–Northeast</td>
</tr>
<tr>
<td>Management similarity</td>
<td>0.28</td>
<td>Estonia–US–Northeast</td>
</tr>
<tr>
<td>Climate similarity</td>
<td>0</td>
<td>Belgium–Denmark</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.197</td>
<td>Austria–US–Southeast</td>
</tr>
</tbody>
</table>

**Table 3.** Posterior means and standard deviations of parameters of the structural model (SM1) for genetic covariances.

<table>
<thead>
<tr>
<th>Parameter(^1)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>51229</td>
<td>809</td>
</tr>
<tr>
<td>$b_1$</td>
<td>68539</td>
<td>1349</td>
</tr>
<tr>
<td>$b_2$</td>
<td>-7173</td>
<td>535</td>
</tr>
</tbody>
</table>

\(^1\mu = \text{Intercept, } b_1 = \text{regression of covariance on genetic similarity, and } b_2 = \text{regression of covariance on management similarity.}\)
past tend to have higher genetic correlations (positive value for $b_1 = 68,392$). The posterior mean of $b_2 (-7173)$ indicated a decrease in genetic covariances when $MS_{ij}$ increased. The definition of management similarity in this study is a contrary indicator: it decreases when milk yields in a pair of environments are more similar. Therefore, a positive relationship between management similarity and genetic covariances is indicated.

The intercept and $b_1$ were estimated precisely, as their posterior coefficients of variation were smaller than 2%; the coefficient of variation of $b_2$ was 7.5%.

Table 4 gives the posterior means of genetic variances of milk yield in the 13 environments using the standard multiple-trait model and the structural model. In general, differences in the posterior means of the genetic variances between both analyses were smaller than or equal to 7%. The largest differences were observed for Israel, Austria, and Belgium, with higher estimates obtained with the structural model. The multiple-trait analysis gave larger posterior means in seven out of 13 comparisons. This suggests a good fit of the structural model, as there were no systematic differences between estimates. Compared with recent estimates used by INTERBULL (1999), calculated from deregressed sire proofs, estimates of genetic variances obtained in this study were lower.

The residual variances (Table 5) were lower with the structural model for genetic covariances in eight of the 13 environments. The estimates of residual variances of milk yield in the five regions in the United States were similar to those found by Rekaya et al. (1999a). As a consequence of the small differences in estimates of genetic and residual variances of milk yield, posterior means of heritabilities in each of the 13 environments using both models were similar (Table 6). The largest heritabilities were for the European countries and Israel. Estimates of heritability of milk yield in the five regions of the United States were very similar, ranging between 0.26 and 0.30.

Genetic correlations for milk yield between the 13 environments using the standard and the structural models were lower than those published by INTERBULL (1999). The genetic correlations ranged between 0.74 (Estonia and Israel) and 0.98 (Midwest and Northwest United States) using the standard multi-
ple-trait model, and between 0.63 (Estonia-Israel) and 0.97 (northwest and southwest United States) using the structural model. Estimates obtained with the latter tended to be slightly lower than those found using the standard multiple-trait analysis. However, some differences were large. Table 7 shows pairs of production environments in which differences in the genetic correlations between the two models were greater or equal to 0.05 (all larger for the standard multiple-trait analysis). The largest differences in genetic correlations were between Estonia and Israel (0.11), and Austria and Estonia (0.06). All pairs of environments with major differences in genetic correlations between the two analyses had very few common sires. As a result of weak genetic ties between these pairs of environments, estimates of the genetic correlations using the standard multiple-trait were very imprecise, in which case point estimates may be of little practical value. This imprecision is illustrated clearly in Table 8, where the 2.5 and 97.5% quartiles of the posterior distributions of genetic correlations are shown as a function of the strength of genetic ties (number of common sires). In the standard multiple-trait analyses, the intervals from 2.5 to 97.5% were much wider than for the structural model when genetic ties were weak (few common sires), and decreased as genetic ties increased. The same pattern was observed with the structural model, but the 2.5 to 97.5% interval was narrower for the same pairs of environments, indicating more precise estimates from the structural model. Even with strong genetic ties (United States-Midwest and United States-Northeast) the genetic correlations were estimated more precisely with the structural model.

The ratio between the DIC obtained from the two models was 7% in favor of the structural model. This indicates a better fit of the structural model, in agreement with the estimated residual variances. The structural model required 15 to 20% more computer time than the standard multiple-trait analysis.

### Table 8. Quartiles (2.5 and 97.5%) of the posterior distribution of genetic correlations between pairs of environments with weak (1, 2, 3), average (4, 5) and strong genetic ties (6).

<table>
<thead>
<tr>
<th>Country or region pair</th>
<th>Standard 2.5%</th>
<th>Standard 97.5%</th>
<th>SM1 2.5%</th>
<th>SM1 97.5%</th>
<th>Common sires</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Estonia–Israel</td>
<td>0.51</td>
<td>0.88</td>
<td>0.54</td>
<td>0.71</td>
<td>0</td>
</tr>
<tr>
<td>2. Israel–Austria</td>
<td>0.53</td>
<td>0.85</td>
<td>0.57</td>
<td>0.75</td>
<td>1</td>
</tr>
<tr>
<td>3. Estonia–Finland</td>
<td>0.63</td>
<td>0.91</td>
<td>0.68</td>
<td>0.86</td>
<td>1</td>
</tr>
<tr>
<td>4. Denmark–Northeast</td>
<td>0.78</td>
<td>0.96</td>
<td>0.85</td>
<td>0.95</td>
<td>100</td>
</tr>
<tr>
<td>5. Belgium–Northeast</td>
<td>0.79</td>
<td>0.96</td>
<td>0.87</td>
<td>0.96</td>
<td>337</td>
</tr>
<tr>
<td>6. Midwest–Northeast</td>
<td>0.85</td>
<td>0.98</td>
<td>0.90</td>
<td>0.98</td>
<td>1275</td>
</tr>
</tbody>
</table>

### CONCLUSIONS

A structural model with three parameters seemed capable of explaining the genetic covariance structure between 13 environments. This model gave a slightly better fit to the data and produced more precise estimates of genetic correlations than a standard multiple-trait model. The performance of the structural model could be improved further if more information regarding management practices and climate conditions were available. In an international genetic evaluation context, it may be difficult to accurately measure climate and management similarity because in some countries little management information is recorded and, further, there is inconsistency between countries in the recording systems. This study illustrates that more detailed global management information can enhance the reliability of international genetic evaluations.

The use of test-day data in national genetic evaluations within some INTERBULL member countries could rapidly increase the number of traits, causing problems in obtaining precise estimates of the needed genetic covariance matrix. A structural model accounting for temporal (time) and spatial (e.g., location, management practices, climate) variation could be a useful tool.

A classical multivariate model may be the best option when adequate information between traits exists in the dataset (e.g., when all animals are recorded for all traits). However, in situations such as international dairy sire evaluations, there is a lack of information for some countries, and a structural model for genetic covariances that makes use of external information could be beneficial.

### 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


Spiegelhalter, D. J., N. G. Best, and B. P. Carlin. 1998. Bayesian deviance, the effective number of parameters and the complexity of arbitrarily complex models. Technical Report, Medical Research Council, Biostatistics Unit, Cambridge, UK.