Multitrait Genetic Evaluation of Jersey Type with Integrated Accounting for Heterogeneous (Co)variances

N. Gengler[1,2], T. Dusseldorf[2], G.R. Wiggans[3], J.R. Wright[3], and T. Druet[1,2]

[1]National Fund for Scientific Research, B-1000 Brussels, Belgium

[2]Animal Science Unit, Gembloux Agricultural University, B-5030 Gembloux, Belgium

[3]Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350, USA

Introduction

In February 1998, a multitrait animal model with canonical transformation was implemented for genetic evaluation of type traits for US Jerseys (Gengler et al., 1999). The method includes multiple diagonalization (Misztal et al., 1995), which is a generalization of canonical transformation to several random effects rather than only additive genetic effects; an expectation-maximization algorithm that permits the use of this approach even if observations for some traits are missing for some cows (Ducrocq and Besbes, 1993); and accounting for inbreeding in the construction of the additive genetic relationship matrix (Wiggans et al., 1995).

Although a common assumption of genetic evaluation models is homogeneity of (co)variances, this assumption is often incorrect across time or herds (e.g., Weigel and Lawlor, 1994). Data can be adjusted to stabilize variances by contemporary group before evaluation. This strategy is used for some yield and type evaluations but is not used for US genetic evaluation of type traits for breeds other than Holstein. Adjustment before evaluation is obviously not optimal. This type of preadjustment is done independently from the evaluation model and, therefore, does not account for genetic or other (co)variances among observations. In addition, preadjustment requires a priori estimation of adjustment factors, which means less flexibility. If a new breed was evaluated or the evaluation model was changed, new adjustment factors would have to be computed. The objective of this study was to develop a strategy for multitrait genetic evaluation of US Jersey type traits that integrates an accounting for heterogeneous (co)variances.

Material and Methods

Current Model

A multitrait (single trait for final score) animal model (Gengler et al., 1999) currently is applied for all traits:

$$\mathbf{y}_t = \mathbf{Xh}_t + \mathbf{Hc}_t + \mathbf{Fd}_t + \mathbf{Ss}_t + \mathbf{Zp}_t + \mathbf{Z^*u}_t + \mathbf{e}_t,$$

where for trait t, $\mathbf{y}$ = vector of type records; $\mathbf{h}$ = vector of fixed effects of herd, date scored, and parity (first or later) group (contemporary group); $\mathbf{c}$ = vector of fixed effects of age group within parity (first or second) and appraisal year group (before 1988 or 1988 and later); $\mathbf{d}$ = vector of fixed effects of lactation stage within parity (first or second) and appraisal year (before 1988 or 1988 and later) group; $\mathbf{s}$ = vector of random effects of interaction of herd and sire; $\mathbf{p}$ = vector of random effects of permanent environment; $\mathbf{u}$ = vector of random additive genetic effects of animals and genetic groups ($\mathbf{u} = \mathbf{a} + \mathbf{Tg}$, where $\mathbf{a}$ = vector of random additive genetic effects of animals expressed as deviations from group means, $\mathbf{g}$ = vector of fixed effects of genetic groups, and $\mathbf{T}$ = incidence matrix that links $\mathbf{g}$ with $\mathbf{u}$); $\mathbf{X}$, $\mathbf{H}$, $\mathbf{F}$, $\mathbf{S}$, $\mathbf{Z}$, and $\mathbf{Z^*}$ = common incidence matrices for all traits that associate $\mathbf{h}$, $\mathbf{c}$, $\mathbf{d}$, $\mathbf{s}$, $\mathbf{p}$, and $\mathbf{u}$, respectively, with $\mathbf{y}$; and $\mathbf{e}$ = vector of random residual effects. Age groups were <25 mo, 25-26 mo, …, 37-38 mo for first parity and <41 mo, 41-42 mo, … 53-54 mo for second parity. Genetic groups were based on birth year (before 1971, 1971-72, … 1991-92, and after 1992). For the remainder of this report, the model will be referred to as $\mathbf{y}_t = \mathbf{Mm}_t + \mathbf{e}_t$.

Applying a canonical transformation based on multiple diagonalization (Misztal et al., 1995) of Var(**s**), Var(**p**), Var(**g**), and Var(**e**) transformed the t observed traits for a given animal i in an environment j (contemporary group) into t unrelated traits ($\mathbf{y}_{Qij}$) with a residual variance of 1 using $\mathbf{y}_{Qij} = \mathbf{Q}\mathbf{y}_{ij}$, where $\mathbf{Q}$ = transformation matrix and $\mathbf{y}_{ij}$ = vector of original traits. If some traits are missing, canonical observations can be obtained from the observed original traits ($\mathbf{y}^{o}_{ij}$) associated with the updated contributions from current solutions on the transformed canonical scales $\mathbf{y}_{Qij} = \mathbf{Q}_1\mathbf{y}^{o}_{ij} + \mathbf{Q}_2\mathbf{M}_j\hat{\mathbf{m}}_{Qij}$ as shown by Ducrocq and Besbes (1993). Then the t mixed-model equation systems are solved based on the general model $\mathbf{y}_{Qij} = \mathbf{M}\mathbf{m}_{Qij} + \mathbf{e}_{Qij}$ and continuous updating for missing records.

Integrated Heterogeneous Variance Adjustment

Meuwissen et al. (1996) developed a method to allow joint estimation of breeding values and heterogeneous variances. Their method was created for milk, fat, and protein yields and is basically a multiplicative mixed model that scales milk production records toward a common phenotypic variance through computation of a heterogeneity parameter each iteration. Then adjustment factors are obtained by modeling those heterogeneity parameters and extracting an expected heterogeneity estimate. This method is appealing because it accounts for (co)variances among observations and heterogeneity factors can be modeled in a flexible manner. However, two major shortcomings are present for application to US Jersey type evaluation. First, the method is univariate, but the US system is multivariate; second, the mean, which has no real meaning for type traits, is scaled. Fortunately both problems can be easily solved.

Multitrait evaluations based on canonical transformation are univariate for the new traits. Using the general heterogeneous variance model proposed by Meuwissen et al. (1996), the following model can be written on the canonical scale:

$$\mathbf{y}_{Qij} = \mathbf{\Gamma}_j(\mathbf{M}\mathbf{m}^{a}_{Qij} + \mathbf{e}^{a}_{Qij}),$$

where $\mathbf{\Gamma}_j = \text{diag}[\exp(\gamma_{jt}/2)]$, which means that all effects are scaled for a given contemporary group j and canonical trait t by $\exp(\gamma_{jt}/2)$ and that the associated variances are scaled by. $\exp(\gamma_{jt})$. Because all associated variances are scaled identically, the hypothesis that the transformation matrix **Q** is still valid and can be accepted.

The problem of the mean can be solved by expressing all original traits as deviations from a general mean. Therefore, if traits on the original scale are not missing, a transformed record that has been adjusted for heterogeneous variance ($\mathbf{y}^{a}_{Qij}$) can be obtained by computing

$$\mathbf{y}^{a}_{Qij} = \mathbf{\Gamma}_j^{-1}\mathbf{Q}(\mathbf{y}^{o}_{ij} - \bar{\mathbf{y}}^{o}) .$$

Similarly, if traits are missing, $\mathbf{y}^{a}_{Qij}$ is obtained by

$$\mathbf{y}^{a}_{Qij} = \mathbf{\Gamma}_j^{-1}[\mathbf{Q}_1(\mathbf{y}^{o}_{ij} - \bar{\mathbf{y}}^{o}) + \mathbf{Q}_2\mathbf{\Gamma}_j\mathbf{M}_j\hat{\mathbf{m}}^{a}_{Qij}].$$

The resulting genetic evaluation method consists of three interdependent iterative systems:

- Solution of regular mixed model equations.
- Update of canonical traits to account for missing original traits.
- Update of adjustment factors for heterogeneous variance.

Mixed-model solution and canonical-trait updates already are part of the current evaluation method.

Update of Heterogeneity Factors

Based on Meuwissen et al. (1996). a heterogeneity parameter z could be developed:

$$z_{jt} = [(\mathbf{y}^{a}_{Qjt})'\mathbf{D}_{jt}\mathbf{e}^{a}_{Qjt} - \sum_{k=1}^{n_j}\lambda_{jtk}]/2,$$

where $\mathbf{D}_{jt} = \text{diag}(\lambda_{jtk})$ = a diagonal matrix with element λ_{jtk} = weight associated with observation k in contemporary group j for trait t. The weight is assumed to be 1 if no original traits are missing and to be <1 if an original trait is missing. Computation of γ_{jtk} follows the methodology proposed in Gengler and Misztal (1996). The variance associated with the heterogeneity parameter is estimated as:

$$\mathrm{Var}(z_{jt}) = [(\hat{\mathbf{m}}_{Qjt})'\mathbf{D}_{jt}\hat{\mathbf{m}}_{Qjt} + 2\sum_{k=1}^{n_j}\lambda_{jtk}]/4.$$

A feature of the method of Meuwissen et al. (1996) is that the modeling of the heterogeneity parameter uses a weighted mixed model on pseudovariates obtained by summing current γ_{jt} with the remaining heterogeneity within contemporary group:

$$(\mathbf{S}'\mathbf{W}_t\mathbf{S} + \boldsymbol{\Lambda}_t^{-1})\boldsymbol{\beta}_t = \mathbf{S}'\mathbf{W}_t[\mathrm{diag}(\gamma_{jt}) + \mathbf{W}_t^{-1}\mathbf{z}_t],$$

where $\boldsymbol{\beta}_t$ = solutions, $\mathbf{S}$ = design matrix linking pseudovariates and $\boldsymbol{\beta}_t$; $\mathbf{W}_t$ = diagonal matrix of iterative weights with $\mathbf{W}_t = \mathrm{diag}[\mathrm{Var}(z_{jt})]$ and $\mathrm{Var}(\boldsymbol{\beta}_t) = \boldsymbol{\Lambda}_t$.

In contrast to Meuwissen et al. (1996), γ_{jt} were scaled towards a common base:

$$\gamma_{jt} = \mathbf{S}\boldsymbol{\beta}_t - \gamma_t^{\mathrm{base}},$$

because mean variances had to be retained for required backsolving. In addition, scaling towards a common base was conceptually similar to the approaches in other studies of type data (e.g., Weigel and Lawlor, 1994; Koots et al., 1994). Definition of the base has no influence on the heterogeneity factor solutions because the approach is similar to an additive base change before and after solving the mixed model equations.

The heterogeneity model can be defined in a general manner. The autoregressive model of Meuwissen et al. (1996) could be considered but was not used. Most studies of type traits applied a structural model (e.g., Weigel and Lawlor, 1994; Koots et al., 1994). The heterogeneity model in this study contained fixed effects to pool information across contemporary groups and an additional random effect that regressed the observed heterogeneity for a given herd-appraisal date back toward the fixed effects. The fixed effects were size of contemporary group and parity (26 classes); mean final score of contemporary group and parity (20 classes); month of classification and parity (24 classes); and 6-mo season, year, and parity (79 classes). This heterogeneity model is a combination of the one used by Koots et al. (1994) for the random effect and the one of Weigel and Lawlor (1994) for fixed effects. This model also pools a priori knowledge and direct observed heterogeneity and, therefore, is conceptually close to the Bayesian approach used for final score of US Holsteins (Weigel and Lawlor, 1994).

Estimation of necessary variance components ideally is done jointly (Meuwissen et al., 1996). However, for this application, variance components were estimated in preliminary studies using Method R (Reverter et al., 1994). For future applications, variance component estimation will be integrated into the system.

Computational Aspects

The publicly available computer program MTJAAM (Gengler et al., 1999) was modified slightly by adding a few lines of code and some restructuring. Estimation of adjustment factors was placed in a subroutine and called from the main program.

Data

The same data used for calculation of official February 2000 US genetic evaluations were used. A total of 563,283 records with a maximum of 16 observed traits from 330,222 cows in 34,402 contemporary groups were included. The pedigree file contained information for 504,211 animals. Solutions from the official February 2000 evaluation were compared with solutions from system to adjust heterogeneous variance.

Results and Discussion

Estimation of Herd-Appraisal Date Variances

Estimates of required herd-appraisal date variances were between 1.6% and 6.9% of total variance. Because the mixed model was weighted according to variance of the heterogeneity factors, the relative weights of the random effects were higher than reflected by those values.

Convergence

Introduction of the heterogeneous variance adjustment slowed down convergence. Despite this, after 200 iteration rounds, convergence expressed as relative squared differences of solutions from one round to the next was around 0.3×10^{-7}. However, optimized solving strategies as discussed by Meuwissen et al. (1996) could improve the convergence rate.

Correlations

Rank correlations among previous solutions and solutions obtained by the new system were high (always near or greater than 0.99). Therefore, as expected, overall ranking of cows with records and of their sires (final score reliability of >0.70) was only minimally affected.

Figure 1 shows the proportion of animals in common based on old and the new rankings for predicted transmitting ability (**PTA**) for final score. Rerankings were most common among the highest ranking animals, especially cows. Only 80% of the top 1% of cows remained in the top 1%. For bulls with a reliability of >0.70 for final score, 90% were in common.

Figure 2 shows that top cows were most affected by the new system. The PTA of all top 100 cows (based on official PTA for final score) were reduced, which indicates that official PTA were biased upwards.

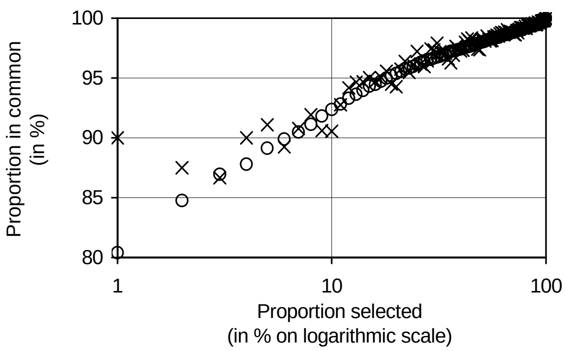

Figure 1. Proportion of animals in common for rankings (top 1% to 100%) of PTA for final score from official and heterogeneous variance-adjusted systems for bulls with a final score reliability of >0.70 (×) and cows (○).

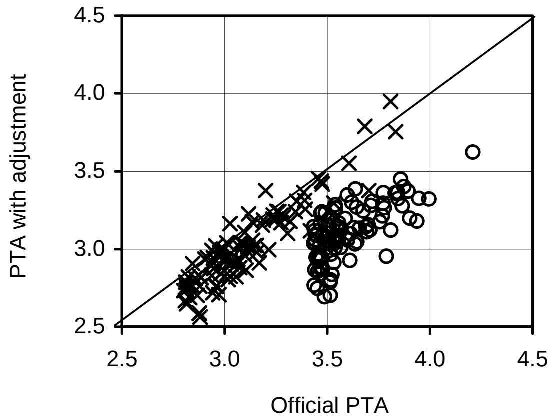

Figure 2. PTA with heterogeneous variance adjustment compared with official PTA for final score for the official top 100 bulls with a final score reliability of >0.70 (×) and for the official top 100 cows (○).

Conclusion

The proposed method for the integration of heterogeneous (co)variance adjustments into the current genetic evaluation system for US Jersey type traits proved to be feasible. The new model is theoretically better than the current one and should give less biased rankings of animals, especially for cows.

Acknowledgments

This study was supported by a research grant from the American Jersey Cattle Association. Nicolas Gengler, who is Chargé de Recherches and Tom Druet, who is Aspirant of the National Fund for Scientific Research, Brussels, Belgium, acknowledge their financial support. Helpful comments from Kent Weigel, University of Wisconsin, are acknowledged.

References

Ducrocq, V., and Besbes, B. 1993. Solution of multiple trait animal models with missing data on some traits. *J. Anim. Breed. Genet. 110, 81-92.*

Gengler, N., and Misztal, I. 1996. Approximation of reliability for multiple-trait animal models with missing data by canonical transformation. *J. Dairy Sci. 79, 317-328.*

Gengler, N., Wiggans, G.R., and Wright, J.R. 1999. Animal model genetic evaluation of type traits for five dairy cattle breeds. *J. Dairy Sci. 82, 1350* (accessed May 24, 2000).

Koots, K.R., Wade, K.M., Kennedy, B.W., Dekkers, J.C.M., Smith, G.C., and Burnside, E.B. 1994. Method and effect of adjustment for heterogeneous variance of Holstein conformation traits. *J. Dairy Sci. 77, 294-302.*

Meuwissen, T.H.E., De Jong, G., and Engel, B. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci. 79, 310-316.*

Misztal, I., Weigel, K., and Lawlor, T.J. 1995. Approximation of estimates of (co)variance components with multiple trait restricted maximum likelihood by multiple diagonalization for more than one random effect. *J. Dairy. Sci. 78, 1862-1872.*

Reverter, A., Golden, B.L., and Bourdon, R.M. 1994. Method R variance components procedure: application on the simple breeding value model. *J. Anim. Sci. 72, 2247-2253.*

Weigel, K.A., and Lawlor, T.J. 1994. Adjustment of heterogeneous variance in genetic evaluations for conformation of United States Holsteins. *J. Dairy Sci. 77, 2563-2571.*

Wiggans, G.R., VanRaden, P.M., and Zuurbier J. 1995. Calculation and use of inbreeding coefficients for genetic evaluation of United States dairy cattle. *J. Dairy Sci. 78, 1584-1590.*