

Estimation of (co)variance components of test day yields for U.S. Holsteins

N. Gengler^{1,2}, A. Tijani¹, G.R. Wiggans³, and J.C. Philpot³

¹Unité de Zootechnie, Faculté universitaire des Sciences agronomiques, B-5030 Gembloux, Belgium;

²Fonds National Belge de la Recherche Scientifique, B-1000 Bruxelles, Belgium;

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

Abstract

(Co)variance components for milk, fat, and protein yields during first lactation were calculated from test-day data from 17,190 cows representing 37 large herds in Pennsylvania and Wisconsin. Initially, four lactation stages of 75 days each were defined, and the test day nearest the midpoint of each interval was used. This approach minimized the number of lactations with missing values. A canonical transformation was used to estimate variance components, which required that all observations with missing values be deleted. Preliminary analysis showed little effect from this selection. Heritabilities usually increased with lactation stage, were highest for milk, and averaged .15. Phenotypic and genetic correlations between milk and protein yields were higher than between milk and fat yields. For each yield trait, the genetic correlation declined from about .90 for adjacent lactation stages to about .75 between lactation stages 1 and 4. When yield traits were from the same lactation stage, the genetic correlation averaged .39 between milk and fat, .78 between milk and protein, and .53 between fat and protein. Phenotypic correlations within lactation stage were $>.90$ between milk and protein and around .65 between milk and fat and between fat and protein. Estimates for the four 75-day lactation stages were extended to provide estimates for twelve 25-day stages (36 traits) using (co)variance functions, which allowed denser coverage of the lactation.

Introduction

One method being investigated for direct use of test-day (TD) data applies a multitrait approach with different TD treated as different traits (Wiggans and Goddard, 1997).

All TD models require estimation of genetic parameters for implementation of an evaluation system. Fortunately, multitrait models can be simplified using canonical transformation, and variance components can be estimated using these uncorrelated traits (Misztal *et al.*, 1995). (Co)variance functions proposed by Kirkpatrick *et al.* (1994) provide a way to extend the estimates of variance components to the whole range of the lactation.

The objective of this study was to estimate variance components for a proposed TD model

using canonical transformation and (co)variance functions.

Materials and Methods

Data

First-lactation records of 17,190 cows from 37 large herds in Wisconsin and Pennsylvania were used. Initially, four lactation stages of 75 days each were defined starting with day 6. For every first lactation started between 1990 and 1996, the TD that was nearest the midpoint of the lactation stage (day 43, 118, 193, or 268) was retained. Only four lactation stages were defined to increase the likelihood of an observation for that stage and separation among stages. In the second part of the study, the four 75-day lactation stages

were extended into twelve 25-day stages. Milk, fat, and protein yields recorded during the four initial lactation stages were the traits analyzed.

Adjustment for lactation curve

Because a TD could occur on any day within a 75-day lactation stage, adjustment for the shape of lactation curves was necessary. Also, because canonical transformation required that the same model apply to all traits and because TD for a particular stage may have different TD effects, effect of TD was removed through a two-stage process. The following model based on Guo and Swalve (1995) was used:

$$y_{ijkmn} = \text{HTDF}_{im} + \text{HAS}_j(b_{1jm}) z_{1n} + \text{HAS}_j(b_{2jm}) z_{2n} + e_{ijkmn} \quad [1]$$

where y_{ijkmn} is a record of cow k in class i for herd, test day, and milking frequency (HTDF) for yield trait m (milk, fat, or protein) and in class j for herd, age, and season (HAS) made at DIM n , b_{1jm} and b_{2jm} are regression coefficients, z_{1n} is $(\text{DIM}_n)^{0.5}$, z_{2n} is $\log(\text{DIM}_n)$, and e is residual effect.

Calving ages for effect of herd, age, and season were defined as 20 to 23, 24 to 25, 26 to 27, and 28 to 35 months; six 2-month calving seasons starting in January were defined. Classes for effect of herd, TD, and milking frequency were required to have at least three observations. If a class for effect of herd, age, and season included <20 TD observations, it was combined with an adjacent class for the same herd and calving age. Effect of calving age and season on persistency was accounted for by nesting regressions within effect for herd, age, and season.

Residuals were considered correlated in the model based on an assumption of constant covariance across all records for yield of cow k . This assumption resulted in a block diagonal structure for $\text{Var}(\mathbf{y})$ and is equivalent to inclusion of a random cow effect with cows considered to be unrelated. The PROC MIXED procedure was used to estimate cow-specific and residual variances with the simplified model.

Residuals from Model [1] were used as adjusted records for further analysis.

Estimation of (co)variance components

The following model was used for analysis across lactation stages:

$$y_{kmpq}^* = \text{HYS}_{mp} + A_{mq} + a_{km} + e_{kmpq}^* \quad [2]$$

where y_{kmpq}^* is the adjusted yield (residual from Model [1]) for yield trait m observed for cow k that calved at age (A) q (in months) in herd-year-season (HYS) p , a is animal effect, and e^* is residual effect for adjusted yield.

Pedigree information was included for animals born in 1982 or later. Variance components were estimated using the expectation maximization REML algorithm described by Misztal et al. (1995).

(Co)variance functions

(Co)variance functions were used to extend the (co)variance components to the whole range of the lactation.

Matrix Σ was defined as a (co)variance matrix for observations of three yield traits in four lactation stages; Φ was defined as the matrix of orthogonal polynomial functions evaluated for milk, fat, and protein during the given lactation stages. Matrix Φ can in this multiple-trait situation be split as: $\Phi = \mathbf{I}_3 \otimes \Phi_t$, where \otimes is the Kronecker (direct) product, \mathbf{I}_3 is an identity matrix of dimension 3 (number of yield traits), and Φ_t is the matrix of orthogonal polynomial functions evaluated at the given lactation stages (Meyer and Hill, 1997). If a full-order fit is assumed, $\Sigma = \Phi \mathbf{K} \Phi'$, where \mathbf{K} is the matrix of covariance function coefficients, which can be estimated as $\mathbf{K} = \Phi^{-1} \Sigma (\Phi')^{-1}$. In a reduced fit case, $\mathbf{K} = (\Phi' \Phi)^{-1} \Phi' \Sigma \Phi (\Phi' \Phi)^{-1}$.

The method is straightforward for the genetic (co)variance matrix \mathbf{G} with $\mathbf{K}_G = \Phi^{-1} \mathbf{G} (\Phi')^{-1}$ but less clear for the residual \mathbf{R} , because \mathbf{R} contains combinations of two types of (co)variances. First, records are affected by nongenetic influences that affect successive observations. The

resulting (co)variance structure would be similar to the genetic (co)variance matrix: decreasing values with increasing distance between TD. Second, records are altered by a random measurement error (Meyer and Hill, 1997). This error can be considered a temporary environmental effect. If one yield trait (e.g., milk) is analyzed, the (co)variance structure of the random measurement error can be assumed to be a diagonal matrix. If correlated yield traits are analyzed (as in this study), this structure becomes more complicated with linked random measurement errors for milk, fat, and protein.

Therefore, for this study, the following decomposition of \mathbf{R} was proposed: $\mathbf{R} = \mathbf{R}^* + \mathbf{E} \otimes \mathbf{I}_4$, where \mathbf{I}_4 is an identity matrix of dimension 4 (number of lactation stages), \mathbf{E} is a 3×3 matrix of temporary environmental variances and covariances among milk, fat, and protein, and \mathbf{R}^* is the part of \mathbf{R} remaining after extracting temporary environmental variances and covariances. This model is simplified because temporary environmental variances and covariances were considered constant over the whole lactation. This simplification is similar to that usually used for residual covariance structures.

The covariance function is then fit to \mathbf{R}^* : $\mathbf{K}_{\mathbf{R}^*} = (\Phi' \Phi)^{-1} \Phi' \mathbf{R}^* \Phi (\Phi' \Phi)^{-1}$. Fitting a full order covariance function for \mathbf{G} , reducing order ($12 - 3 = 9$) of the covariance function for \mathbf{R}^* , and estimating \mathbf{E} results in a model equivalent to full fit (Meyer and Hill, 1997). A reduced fit model also was necessary for \mathbf{R}^* , because a full fit on \mathbf{R} yielded unlikely negative residual (co)variances.

Kirkpatrick *et al.* (1994) proposed a method to separate \mathbf{R}^* and \mathbf{E} based on fitting a covariance function on the covariances only and estimating the variances without the measurement error. However, this approach was too difficult to use in this study. Therefore, the following strategy was implemented. Matrices \mathbf{G} and \mathbf{R}^* were assumed to have similar rates of decrease of covariances over time with increasing distance between TD. Estimation of \mathbf{R}^* was based on regression of elements of \mathbf{R} on \mathbf{G} with replacement of critical elements with their estimates. Then \mathbf{E} was estimated as the remaining part of \mathbf{R} that is not contained in \mathbf{R}^* .

Results and Discussion

(Co)variance matrices were estimated across lactation stages. Genetic and residual variances are in Table 1. Variances tended to increase with increasing lactation stage. Genetic and residual variances were highest for final lactation stage except for fat residual variance, which was highest at the start of lactation.

Heritabilities, genetic correlations, and phenotypic correlations are in Tables 2 and 3. Table 2 shows the results from analysis of the original data set of four 75-day stages. Table 3 shows the results from the proposed (co)variance functions applied to the 36 traits defined by subdividing each 75-day lactation stage into three 25-day stages. All heritability estimates in Table 2 were between .10 and .18. Estimated heritabilities usually were higher for milk yield than for protein and fat yields. Heritability estimates tended to increase with increased lactation stage for all three traits

Genetic and phenotypic correlations in Table 2 are all positive and between .22 and .98. Genetic correlations were high (as expected) between lactation stages for the same yield trait. Genetic correlations among yield traits within the same lactation stage were lower between milk and fat and between fat and protein than between milk and protein. Genetic correlations between traits at the same stage were lower for stages 2 and 3 than for stages 1 and 4.

Similar to genetic correlations among yield traits within the same lactation stage, phenotypic correlations were lower between milk and fat and between fat and protein than between milk and protein. These phenotypic correlations also were particularly high (>.9 between milk and protein within each lactation stage).

(Co)variance function results are in Table 3. Values for the middle 25-day lactation stage of each 75-day stage were quite similar to those in Table 2. Results showed that the (co)variance functions used were able to extrapolate correctly between known lactation stages and provided reasonable estimates of the covariance structure at the ends of the lactation curve. Additional reduction in parameters may be possible through

further investigation.

Conclusions

The estimation of (co)variance components is crucial for genetic evaluation systems using TD yields. The present study used a simple approach based on canonical transformation to construct (co)variance matrices. The first results from the four 75-day stages provided estimates for genetic and residual (co)variance for three yield traits across four lactation stages.

Milk and protein yields were found to be more highly correlated than milk and fat yields. For each yield trait, the genetic correlation declined from about .90 for adjacent lactation stages to about .75 between lactation stages 1 and 4. When yield traits were from the same lactation stage, the genetic correlation averaged .39 between milk and fat, .78 between milk and protein, and .53 between fat and protein. Phenotypic correlations within lactation stage were $>.90$ between milk and protein and around .65 between milk and fat and between fat and protein.

Results from the (co)variance functions provided estimates for 36 traits. Results corresponding to the original 75-day stages were comparable with those generated by the functions. Other estimates (extrapolations or interpolations) illustrated the validity of the proposed (co)variance function methodology as

a way to increase the density of estimates over the whole lactation.

References

- Guo, Z., and Swalve, H.H. 1995. Modeling of the lactation curve as a sub-model in the evaluation of test day records. Proc. Open Session INTERBULL Annu. Mtg., Prague, Czech Republic, Sept. 7-8, 1995. Int. Bull. Eval. Serv. Bull. No. 11. Dep. Anim. Breed. Genet., SLU, Uppsala, Sweden.
- Kirkpatrick, M., Hill, W.G., and Thompson, R. 1994. Estimating the covariance structure of traits during growth and aging, illustrated with lactation in dairy cattle. Genet. Res. Camb. 64:57-66.
- Meyer, K., and Hill, W.G. 1997. Estimation of genetic and phenotypic covariance functions for longitudinal or "repeated" records by restricted maximum likelihood. Livest. Prod. Sci. 47:185-200.
- 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.
- Wiggans, G.R., and Goddard, M.E. 1997. A computationally feasible test day model for genetic evaluation of yield traits in the United States. J. Dairy Sci. 80:1795-1800.

TABLE 1. Genetic and residual variances for milk, fat, and protein yields across four lactation stages.

Yield trait	Midpoint (DIM) of lactation stage	Genetic variance	Residual variance
Milk, kg ² × 100	43	361	2129
	118	377	1933
	193	412	1948
	268	484	2263
Fat, g ²	43	4850	42,975
	118	5951	37,182
	193	5781	33,967
	268	6557	37,123
Protein, g ²	43	2694	17,977
	118	2458	17,688
	193	3144	18,876
	268	3931	22,682

TABLE 2. Heritabilities (on diagonal), genetic correlations (above diagonal), and phenotypic correlations (below diagonal) for milk, fat, and protein yields across four lactation stages.

Yield trait	Midpoint (DIM) of lactation stage	Milk yield at midpoint (DIM) of lactation stage				Fat yield at midpoint (DIM) of lactation stage				Protein yield at midpoint (DIM) of lactation stage			
		43	118	193	268	43	118	193	268	43	118	193	268
Milk	43	.15	.91	.82	.74	.51	.31	.30	.36	.80	.76	.74	.71
	118	.56	.16	.98	.90	.34	.26	.27	.35	.63	.72	.74	.74
	193	.48	.63	.17	.96	.28	.24	.29	.41	.54	.70	.76	.79
	268	.39	.52	.62	.18	.26	.24	.31	.49	.48	.65	.73	.83
Fat	43	.64	.34	.29	.22	.10	.88	.84	.77	.63	.59	.61	.57
	118	.28	.57	.33	.28	.38	.14	.96	.88	.36	.36	.40	.41
	193	.27	.36	.63	.36	.34	.47	.15	.96	.35	.41	.49	.53
	268	.25	.32	.41	.72	.28	.39	.49	.15	.35	.43	.54	.64
Protein	43	.91	.45	.38	.30	.64	.26	.24	.22	.13	.93	.84	.73
	118	.48	.90	.53	.42	.33	.58	.35	.32	.47	.12	.97	.89
	193	.42	.55	.91	.54	.31	.34	.65	.43	.40	.56	.14	.96
	268	.36	.46	.56	.93	.25	.30	.39	.76	.33	.45	.57	.15

