

Estimation of Heritability for Yield of U.S. Dairy Cattle

C.P. Van Tassell, G.R. Wiggans, H.D. Norman, and R.L. Powell

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

Abstract

Heritabilities (h^2) for milk, fat, and protein yields were estimated from first lactation data used for USDA-DHIA genetic evaluations. Contemporary group assignments and standard deviations (**SD**) within herd-year were determined with the procedure used for national evaluations. Pedigree data were included for animals born since 1970; yield data were included for cows born since 1980. Lactation records were divided into four mutually exclusive data sets based on SD. Ranges for SD were chosen so that data sets were approximately equal in size. Method R was used to estimate h^2 with 25 different random samples of half of the data for each data set. Because of the large number of Holstein observations, estimates of h^2 for Holsteins were based on random subsets of the complete data file; each subset including approximately 5% of the data. Mean h^2 estimates increased with SD, and estimates ranged from .18 to .51 across breeds. Repeatability estimates for milk yield of Holsteins were approximately .50 and did not change with SD. Previous h^2 used for USDA-DHIA genetic evaluations, which averaged .25 and ranged from .20 to .30, appeared to be too low. Based on these results and validation of proposed changes, a mean h^2 of .30 with a range of .25 to .35 is used for current USDA-DHIA evaluations.

Introduction

Estimates of variance components are required by USDA's Animal Improvement Programs Laboratory (**AIPL**) to calculate national genetic evaluations of dairy cattle. However, because of computational limitations, parameters have never been estimated using the complete national data set of lactation records maintained at AIPL.

Many researchers have estimated heritabilities (h^2) higher than .25 for yield traits, and most countries now assume h^2 higher than .25. The objective of this study was to estimate parameters using the same data and analysis model as are used for genetic prediction. Additionally, validation of the improvement in predicted transmitting abilities (**PTA**) was desired.

Materials and Methods

Data

Data were obtained from the AIPL database for the five major breeds of dairy cattle: Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey. Data for Red and Whites were combined with Holstein data. The AIPL database includes historical yield and pedigree information for cows enrolled in Dairy Herd Improvement programs throughout the United States.

The complete data set included records from first lactations of all cows born after 1980. This data set was reduced by removing noninformative records. The remaining data set was divided into four mutually exclusive groups based

on standard deviation (**SD**) within herd-year. These data sets formed by dividing the complete data set were called quartile sets. Quartile set 1 had records with the lowest SD; quartile set 4 had records with the highest SD.

Pedigree data were included for animals born since 1970. Pedigree data were reduced by using an iterative process of removing records of parents or progeny without observations if they did not contribute genetic ties between animals with observations.

Additional analyses of Holstein records for milk yield were conducted to determine the influence of time and herd-year SD on h^2 estimates. These effects are partially confounded in the complete data set because of a trend of increasing SD over time. Two subsets of first lactation data were created based on birth year of cow: 1970 to 1979 and 1980 to 1989. Pedigree information for the two data sets was included for animals born since 1960 and 1970, respectively. The quartile divisions from the complete Holstein data set were used for these subsets so that the influence of time could be assessed.

Parameter estimation

Method R (Reverter *et al.*, 1994), a relatively new procedure that allows analysis of large data sets, was used for parameter estimation. Method R requires R values, which are regressions of predicted random effects calculated using “complete” data on predicted random effects calculated using random subsets of the same data. All R values are 1 if the parameters are appropriate for the population.

The main advantage of using Method R is that large data sets can be used for parameter estimation because the procedure is based on repeated solutions of standard mixed model equations. Estimation of parameters for populations with as many as 4 million animals has been accomplished using Method R (Misztal, 1997).

The model used for the complete data analysis was

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e},$$

where \mathbf{y} , $\boldsymbol{\beta}$, \mathbf{u} , and \mathbf{e} are vectors of observations, fixed effects, random effects, and random residual effects, respectively, and \mathbf{X} and \mathbf{Z} are incidence matrices. For this analysis, $\boldsymbol{\beta}$ included contemporary groups in a herd, and \mathbf{u} included animal genetic effects. Standard assumptions were made about parameter means and variances:

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}; \text{Var} \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} & \mathbf{Z}\mathbf{G} & \mathbf{R} \\ \mathbf{G}\mathbf{Z}' & \mathbf{G} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{bmatrix}.$$

where $\mathbf{R} = \mathbf{I}\sigma_e^2$, $\mathbf{G} = \mathbf{A}\sigma_a^2$, and \mathbf{A} is the numerator relationship matrix describing genetic relationships among animals.

Solutions to the equations were obtained using ITPACK (Kincaid *et al.*, 1984; Kincaid *et al.*, 1996) applied to Henderson's (1984) mixed model equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\boldsymbol{\gamma} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix},$$

where $\boldsymbol{\gamma} = \sigma_e^2/\sigma_a^2 = (1-h^2)/h^2$.

Data subsets were chosen randomly with a chance of .5 that any observation would be included. Data in the subsets were adjusted for solutions of fixed effects from the complete data set prior to analysis; i.e., subset data (\mathbf{y}_s) were randomly selected from $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. Solutions for random effects (\mathbf{u}_s) for \mathbf{y}_s were obtained as solutions to

$$[\mathbf{Z}'_s\mathbf{Z}_s + \mathbf{G}^{-1}\boldsymbol{\gamma}]\hat{\mathbf{u}}_s = [\mathbf{Z}'_s\mathbf{y}_s],$$

where \mathbf{Z}_s is the incidence matrix relating animal effects to randomly selected observations.

Finally, the regression of estimates of random effects from complete data on estimates of those same effects from data subsets was calculated:

$$\text{R value} = \frac{\hat{\mathbf{u}}' \mathbf{A}^{-1} \hat{\mathbf{u}}_s}{\hat{\mathbf{u}}_s' \mathbf{A}^{-1} \hat{\mathbf{u}}_s}.$$

Estimated h^2 was adjusted until the R value was 1.

Estimates of repeatability (fraction of variance accounted for by genetic, herd-sire interaction, and permanent environmental effects) was estimated using MTDFREML (Boldman *et al.*, 1995). This analysis assumed cows as the only random source of variation and fit contemporary groups as fixed effects. Cow effects were assumed to be uncorrelated. Fraction of variance explained by cows in this model provides an estimate of repeatability.

Validation of estimates

Effect of changes in the genetic evaluation system on bull evaluations as they progress to include second-crop daughter data also was examined. Differences between PTA for protein based on Holstein data used for May 1997 USDA-DHIA evaluations and PTA based on data from cows calving before January 1, 1993, were examined for the proposed and previous evaluation systems. Evaluations based on first-crop daughters had to include data from ≥ 10 but ≤ 500 daughters; evaluations based on first- and second-crop daughters had to have an increase in reliability of $\geq .09$ between the two evaluations. A total of 263 Holstein bulls met these criteria.

Results and Discussion

Estimated h^2 ranged from .18 to .51 (.26 to .48 for Holsteins and Jerseys) (Table 1). In general, h^2 estimates increased with herd-year SD. Trait and breed combinations without strong evidence of increasing trend included Guernsey fat and protein yields and Holstein fat yield. Mean h^2 estimates by quartile set were .29, .33, .39, and .46 for milk yield; .30, .33, .33, and .37 for fat yield; and .25, .32, .30, and .35 for protein yield.

Estimated h^2 and standard errors of estimates for milk yield of Holsteins for data split by time were compared with results from the complete data set (Table 2). Because the estimates showed no evidence of time differences, h^2 changes over time can be considered only as a function of herd-year SD; evidence for additional time trend was not found.

Repeatability estimates were calculated only for milk yield of Holsteins. Estimates were .495, .499, .505, and .494 by quartile set. Approximate standard errors were $\leq .006$. Thus, no evidence exists for increasing h^2 to be associated with reduced residual variance. The change must be associated with other effects in the model; h^2 increases must correspond to decreases in fraction of variance associated with herd-sire interaction or permanent environmental effects.

Based on these results, the following changes to the USDA-DHIA genetic evaluation system were proposed: 1) increase mean h^2 from .25 to .30; 2) change range of h^2 from .20-.30 to .25-.35; 3) reduce fraction of variance due to herd-sire interaction from .14 to .10 and due to permanent environment from .16 to .15; and 4) impose a floor and ceiling of four phenotypic SD on yield deviations of animals from contemporary group.

Improvement in mean changes of consecutive evaluations and increased correlation of PTA from consecutive evaluations should result if changes proposed were appropriate for the analysis system. Validation results from analysis of this proposal showed evidence of improvement: reduced mean changes and increased correlation of the two sets of PTA. This system is currently used for USDA-DHIA genetic evaluations.

An alternative proposal was evaluated with the following differences from the original proposal: 1) increase mean h^2 from .25 to .30 and allow mean h^2 to vary by birth year with mean of .30 used for animals born in 1990; and 2) increase range of h^2 from .20-.30 to .20-.40. However, validation of this alternative proposal did not indicate an improvement. Although correlations of PTA from the original and proposed systems were $>.99$ for all comparisons made, mean changes in PTA increased, and correlations from the early and later evaluations were nearly equal for both systems.

Summary

Estimated h^2 using Method R provided evidence for increasing the h^2 for USDA-DHIA genetic evaluations. Proposed system changes

were tested, and prediction was improved by increasing h^2 and yield deviation limits.

Literature Cited

- Boldman, K.G., L.A. Kriese, L.D. Van Vleck, C.P. Van Tassell, and S.D. Kachman. 1995. A Manual for Use of MTDFREML. A Set of Programs To Obtain Estimates of Variances and Covariances [draft]. USDA, Agricultural Research Service.
- Henderson, C.R. 1984. Application of Linear Models in Animal Breeding. University of Guelph Press, Guelph, ON, Canada.
- Kincaid, D.R., T.C. Oppe, J.R. Respass, and D.M. Young. 1984. ITPACK 2C Users' Guide. CNA-191, Center for Numerical Analysis, University of Texas, Austin.
- Kincaid, D.R., J.R. Respass, D.M. Young, and R.G. Grimes. 1996. ITPACK 2C: a FORTRAN package for solving large sparse linear systems by adaptive accelerated iterative methods. [HTTP://www.netlib.org/itpack/user2c.tex](http://www.netlib.org/itpack/user2c.tex).
- Misztal, I. 1997. Estimation of variance components with large-scale dominance models. *J. Dairy Sci.* 80:965-974.
- Reverter, A., B.L. Golden, R.M. Bourdon, and J.S. Brinks. 1994. Method R variance component procedure: application on the simple breeding value model. *J. Anim. Sci.* 72: 2247-2253.

TABLE 1. Means of heritability estimates for first lactation milk, fat, and protein yields with maximum estimated standard deviations (SD) for four quartile sets by breed of dairy cattle from 25 Method R samples.

Breed	Yield trait	Maximum SD	Quartile set			
			1	2	3	4
Ayrshire	Milk	.06	.25	.24	.33	.44
	Fat	.07	.26	.27	.31	.36
	Protein	.06	.27	.29	.27	.36
Brown Swiss	Milk	.06	.27	.34	.37	.48
	Fat	.08	.28	.37	.37	.43
	Protein	.08	.23	.37	.39	.43
Guernsey	Milk	.05	.23	.35	.45	.51
	Fat	.04	.31	.31	.31	.30
	Protein	.04	.18	.32	.18	.18
Holstein	Milk	.03	.31	.33	.35	.41
	Fat	.03	.36	.33	.33	.38
	Protein	.04	.26	.28	.29	.36
Jersey	Milk	.02	.38	.40	.45	.48
	Fat	.02	.31	.35	.35	.38
	Protein	.02	.31	.32	.37	.42

TABLE 2. Means of h^2 estimates for first lactation milk yield for four data sets with estimated standard deviation (in parentheses) for four quartile sets for Holsteins from 25 Method R samples.

Analysis data ¹	Quartile set			
	1	2	3	4
Complete (1980-)	.31 (.02)	.33 (.03)	.35 (.03)	.41 (.03)
Early (1970-79)	.33 (.03)	.33 (.06)	.31 (.05)	.35 (.07)
Late (1980-89)	.33 (.03)	.34 (.03)	.35 (.04)	.41 (.03)

¹Dates in parentheses indicate birth years for cows with first lactations included in analysis.