

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Papers and Publications in Animal
Science

Animal Science Department

March 1987

Genetic Parameters for First Lactation Milk Yields at Three Levels of Herd Production

J. C. De Veer
Cornell University

L. Dale Van Vleck
University of Nebraska-Lincoln, dvan-vleck1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Animal Sciences Commons](#)

De Veer, J. C. and Van Vleck, L. Dale, "Genetic Parameters for First Lactation Milk Yields at Three Levels of Herd Production" (1987). *Faculty Papers and Publications in Animal Science*. 302.

<https://digitalcommons.unl.edu/animalscifacpub/302>

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Genetic Parameters for First Lactation Milk Yields at Three Levels of Herd Production

J. C. De VEER and L. D. VAN VLECK

Department of Animal Science

Cornell University

Ithaca, NY 14853

ABSTRACT

A multivariate linear model was used to estimate sire variance and covariance components and residual variance components for first lactation milk yield and logarithms of yield at three herd production levels using Restricted Maximum Likelihood with the Expectation-Maximization algorithm. Data for four separate analyses were 305-d, mature equivalent first lactation milk records from cows sired artificially in the north-eastern United States that freshened in 1970, 1971, 1976, and 1984. Respective numbers of records for each year were 42,618, 40,207, 33,581, and 34,196. Corresponding numbers of sires were 298, 289, 305, and 313. Herd production level was defined by mean yield of all cows freshening in same herd-year-season.

For untransformed records sire and residual components of variance increased as mean increased, both within and between years. Correlations between sire effects at different production levels were all above .85. Heritabilities increased as production level increased. These results indicate that it may be necessary to account for heterogeneous genetic and environmental variance in sire evaluations.

For logarithms of yield, sire components of variance were similar for each of the three production levels within a year. Residual components for logarithms decreased as production level increased. Change in variance from one production level to another was considerably more for logarithms than for untransformed yields.

INTRODUCTION

Milk yield has been long recognized to show important "scale effects," i.e., the standard deviation increases with the mean (11). Genetic components of variance also are related to the mean although this relationship has not been consistent [see, e.g., (2, 3, 11, 14, 17, 18, 20, 23)].

An evaluation procedure using a linear model with common variance assumptions for the logarithms of records implies that the majority of the model effects are multiplicative. Some evidence for this was given by Everett and Keown (4).

Mirande (16, 17) obtained estimates of sire and residual variance taking records from different years and production levels as independent populations. Herd production levels were assigned by rolling herd averages. Method of analysis did not account for relationships between sires nor bias due to selection of the parents of the sires, which prevents interpretation of temporal changes in the estimates. Nevertheless at a given time, sire and residual components of variance increased as the production level and heritability also increased. However, for records transformed to logarithms, the residual variance was largest at low production and smallest at high production levels. This result can be mimicked by simulation of records using a multiplicative effects model to which different means are added to determine the level of each record, but the result does not prove nor disprove which model is appropriate for raw data or logarithms.

Hill et al. (11) split production records from British Friesian cows into two levels on the basis of mean milk yield of the herds. Variance components were larger for the high production level. For the logarithm of records, the within sire families component of variance was slightly smaller for the high level than for the low level. For data split on the basis of residual sum of squares rather than on means, the variance components increased markedly with level on

Received October 14, 1986.

Accepted March 12, 1987.

both the linear and logarithm scales. Heritabilities were generally larger for logarithms and larger for the high level. Change in heritability associated with level was larger for logarithms both for data split on means and on variance. Estimates of genetic correlation between levels were not different from 1, indicating absence of genotype-environment interaction.

The purpose of this study was to estimate variance and covariance components for milk yield and for the natural logarithm of milk yield using a multivariate model in which yields at three herd production levels are considered three correlated variables; i.e., levels play the role of traits in terms of the usual multiple trait analysis. Relationships among sires were to be considered. The primary goal was to estimate the genetic correlations between genetic expression at the three production levels. The patterns for heritability and variances have already been well established (3, 4, 11, 17, 20, 23).

Genetic correlations provide information about genotype-environment interaction and can be used to quantify loss of information from using records made at a level different from that where the selected animals will perform.

DATA

The original data set consisted of 1,129,883 305-d, mature equivalent first lactation milk records of Holstein cows sired by AI bulls in the northeastern United States. Records were coded in the same herd-year-season (HYS) if lactations started in the same herd and season. Two seasons were defined per year, December through April and May through November.

Total number of sires with daughters having records in the data set was 5582. To calculate all relationships among sires, additional sires were used that did not have daughters in the file but that were related to sires with daughters. Total number of sires in the relationship matrix was 6555. Of these, 1544 appear in pedigrees of other sires.

A subset of sires was selected for each analysis because computer time and size required to include more than 350 sires in each analysis were prohibitive. The only informative

HYS are those including daughters of more than one of the involved sires. Thus, number of records actually used in the analysis was less than total number of daughters in the subset of bulls. To reduce this loss of records and because of the important time trend observed in mean yields [e.g., (16)], subsets were chosen by selecting short intervals for freshening dates such that about 300 sires would be included in each analysis. The original intent was to use only records of sampling daughters to reduce the effect of selection on the sire variance. The reduction, however, in number of sire comparisons within HYS due to lack of contemporaries was drastic, and therefore, daughters of all sires in the subset were included, discarding only those from HYS in which only one sire in the subset was represented. A basic goal was to estimate the genetic correlations from sire values in different production levels. Selection was not thought to affect the estimate of the genetic correlations.

Definition of Herd Production Levels

Herd production level is the result of a combination of factors. These factors are mostly unknown, and only a few can be identified. For lack of a better procedure, each record was assigned to a herd production level using the average milk yield of all first lactations initiated in the same HYS, regardless of inclusion or elimination of the record from analysis of the subset. In preliminary analyses, assignment to levels by phenotypic averages of all cows in the HYS resulted in similar estimates of variance and covariance components.

To include as many sires as possible in each analysis under restrictions imposed by computer storage, only three levels were recognized, i.e., low, medium, and high. Upper and lower limits defined each level with records discarded that were not within any level.

Table 1 contains limits used in defining levels for each group of data. Levels for year 1984 could not be defined to match the same limits used for 1970, 1971, and 1976, because an approximately even distribution of the records in the levels could not be obtained, i.e., too few records were at the low level because overall yield had increased.

TABLE 1. Upper and lower limits (kg) used to classify records into levels on the basis of average yield of all first lactations in the herd-year-season.

Data set	Low level		Medium level		High level	
	Lower	Upper	Lower	Upper	Lower	Upper
1970	4086	6129	6356	7718	7945	9988
1971	4086	6129	6356	7718	7945	9988
1976	4086	6129	6356	7718	7945	9988
1984	5448	7264	7264	8172	8172	9988

Model

The following model was used, both for records and logarithms of records.

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}$$

where y_i = vector of n_i records (or logarithms) made at level i ;

X_i = matrix indicating HYS effects in level i associated with the records in y_i ;

β_i = vector of fixed effects of HYS for level i ;

Z_i = matrix indicating sires of cows associated with records in y_i (contains null columns corresponding to sires with no daughters at level i);

s_i = vector of transmitting abilities of all sires in the analysis for level i , where the sire effects may differ from one production level to another;

e_i = vector of residual random effects associated with records in y_i .

Also:

$$E(y_i) = X_i\beta_i \quad E(s_i) = 0 \quad E(e_i) = 0 \quad \text{Cov}(s,e) = 0$$

$$\text{Var} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{bmatrix} * A = S * A$$

where S is a square matrix of order 3 containing variances of and covariances among the sire effects at the three levels, A is the relationship matrix among the sires, and the symbol, $*$, stands for direct (Kronecker) product.

$$\text{Var} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \begin{bmatrix} e_{11}I_{n_1} & 0 & 0 \\ 0 & e_{22}I_{n_2} & 0 \\ 0 & 0 & e_{33}I_{n_3} \end{bmatrix}$$

where e_{ij} represents residual variance common to all records at level i , and I_{n_i} is an identity matrix of order n_i , which is the number of records at level i . Matrix A was computed using the algorithm of Hudson et al. (13).

The estimation procedure was described by Henderson (8, 9) as an expectation-maximization-type algorithm for restricted maximum likelihood (REML) that is similar to iterated minimum variance quadratic unbiased estimation except that at each round of iteration, expectations are taken pretending that $G = \hat{G}$ and $R = \hat{R}$, where \hat{G} and \hat{R} were estimates from the previous round of iteration. The model used in this study corresponded to that in Schaeffer et al. (21).

RESULTS AND DISCUSSION

Two types of results must be recognized: those relative to the assumed base population (years 1970 and 1971) and those for later records (second half of 1976 and of 1984). The following notation refers to the subsets.

Subset	Freshening dates	Scale
Y70	1/1/70 to 12/31/70	Linear
Y70-L	1/1/70 to 12/31/70	Log
Y71	1/1/71 to 12/31/71	Linear
Y71-L	1/1/71 to 12/31/71	Log
Y76	6/1/76 to 12/31/76	Linear
Y76-L	6/1/76 to 12/31/76	Log
Y84	6/1/84 to 12/31/84	Linear
Y84-L	6/1/84 to 12/31/84	Log

TABLE 2. Number of records by level of production, herd-year-seasons, and sires included in each year.

Year	Level of production			Herd-year-seasons	Sires
	Low	Medium	High		
1970	9740	27,163	5715	7478	298
1971	9699	25,335	5173	6960	289
1976	6845	18,660	8076	6078	305
1984	8683	10,945	14,568	4987	313

In some analyses, solutions reached a small rate of change from one round of iteration to the next after about 10 rounds. In others, however, solutions reached a point of little change but then suddenly diverged, leading to huge sire components of variance and covariance within two or three rounds. The covariance matrix of sire effects was still positive definite, but environmental components derived from the solutions became negative and heritabilities much larger than unity. Because solutions before divergence were about as expected and observed in other analyses, it was suspected they were near REML estimates. To support this conjecture, the same program was run but with solutions for covariances restricted to zero. Thus, the analysis became a set of three independent analyses. Convergence was obtained for sire and residual components of variance at all three levels. Both sets of results are reported.

Table 2 contains number of records, sires, and HYS included in each subset. In all analyses, more than 90% of the sires had daughters in all three levels. Means for yield and logarithm of yield corresponding to each data set are in Table 3.

Date when sires of each subset entered

service is a time reference for the population being sampled. Median years entering service were 1965, 1966, 1969, and 1975 for the corresponding data subsets of 1970, 1971, 1976, and 1984. Estimated variance and covariance components are listed in Table 4. Subsets Y70-L and Y71-L seemed to converge until the round shown in the table, after which solutions for sire components started to increase at an increasing rate, accompanied by a greater rate of reduction in residual components. No explanation could be found for this phenomenon. Changing starting values did not change the values to which the solutions seemed to approach, and divergence still occurred. Variance estimates before divergence seemed to agree with variance components obtained by Miranda (16) from essentially the same data set using Henderson's method 3 for a sire model but not considering relationships.

Subsets that did not converge were analyzed with the same program under the restriction of zero covariances among sire effects at different levels. With the restriction, convergence was obtained. Estimates obtained in these cases were not very different from their respective counterparts for the unrestricted analyses just prior to divergence. This agreement might

TABLE 3. Means of all records in each level for all data sets.

Year	Mean yield			Mean yield		
	Low	Medium	High	Low	Medium	High
	(kg)			(1000 log kg)		
1970	5543	6979	8343	8595	8834	9017
1971	5562	7007	8354	8600	8838	9018
1976	5503	7068	8528	8590	8848	9038
1984	6595	7699	8785	8773	8932	9065

TABLE 4. Estimated variance and covariance components (kg² in linear scale, [1000 × log kg]² in logarithmic scale) for low, medium, and high production levels.

Analysis	No. of round	Sire component of variance			Sire component of covariance			Residual component of variance		
		Low	Medium	High	Low to medium	Low to high	Medium to high	Low	Medium	High
Linear scale:										
Y-70	14	39,030	62,853	100,457	48,076	53,894	69,225	1,193,221	1,328,419	1,492,072
Y-71	25	45,278	66,215	133,598	53,297	73,193	89,480	1,158,063	1,378,899	1,534,178
Y-76 ¹	16	32,479	60,536	85,547	1,042,611	1,321,872	1,631,454
Y-84 ¹	36	75,521	97,356	150,908	1,501,468	1,827,214	2,092,669
Log scale:										
Y-70-L ²	13	2508	2345	2415	2201	2103	2047	46,948	31,352	23,466
Y-71-L ²	11	2875	2322	2618	2371	2488	2232	46,105	32,558	24,809
Y-70-L ¹	18	2211	2440	2309	47,090	31,349	23,407
Y-71-L ¹	12	2880	2498	2070	46,132	32,544	24,919
Y-76-L ¹	15	1166	1228	1273	41,167	31,327	26,722
Y-84-L ¹	20	1594	1799	1984	43,881	38,420	32,478

¹ Sire covariances restricted to equal zero (independent analyses for each level).

² Values just prior to divergence.

indicate that estimates obtained just prior to divergence in the covariance model are approximately REML estimates.

Analyses of early data sets indicated that the residual component of variance is not homogeneous for the different levels. Log transformation did not reduce heterogeneity of residual variances but, on a proportional basis, made it considerably greater as shown in Figure 1 in which estimated components for year 1970 are plotted by level for linear and log scales. Similar patterns were observed in all other analyses. Sire components of variance (Figure 2), however vary less from one level to another on log transformed data than on the original scale. In Hill et al. (11) with records divided in two groups the variance within sire families on linear scale increased as the mean increased but was practically constant on the log scale. Heritability estimates in that study were higher for high production level on both scales. Increase in heritability from low to high level was larger for the log scale (11) although not in the present study.

Estimates of heritabilities and genetic correlations are in Table 5. As the level increased, heritabilities increased. In the early years, heritabilities were larger for logarithms than for untransformed records but were smaller in later years. Miranda and Van Vleck (17) reported heritabilities averaged essentially the same for both scales.

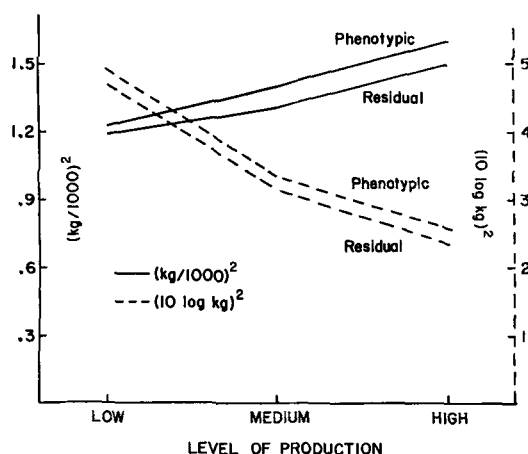


Figure 1. Estimates of variance components by level of production for 1970 on linear and logarithmic scales.

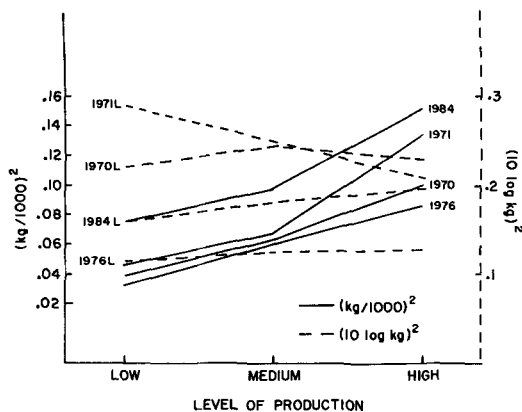


Figure 2. Estimates of sire variance components by level of production for each year on linear and logarithmic scales.

Estimated correlations among sire values at different levels were large in all cases (all .85 or greater) but were slightly smaller for logarithms than for their untransformed counterparts, which may be due to convergence problems that limited the number of iterations that could be done on the log data. Danell (3) and Syrstad (22) in reviews have reported estimates of genetic correlations from other studies ranging from .80 to 1.00.

Time Effect on Variance Components

The population of sire transmitting abilities sampled in this study has undergone genetic selection estimated to be responsible for an average increase of 18 kg of milk per lactation/kg (12). Phenotypic mean yield in this data set has changed at a rate of about 90 kg/yr. This increase in mean is associated with an increase in variance components, which indicates some factors interact by multiplying their effects. Variance may also increase because of a wider range of factors at a given time and place; e.g., different genetic populations of sires or more efficient techniques for deliberate differential treatment of cows. Variance may also increase as herds become larger. However, genetic variability usually is thought to decrease as a result of genetic selection, and environmental variability may decrease due to more standardized treatment of all cows receiving the same amount of feed or may increase if cows in

TABLE 5. Estimated heritability and genetic correlations for low, medium, and high production herds.

Analysis	Heritability			Genetic correlations		
	Low	Medium	High	Low to medium	Low to high	Medium to high
Linear scale:						
Y-70	.13	.18	.25	.97	.86	.87
Y-71	.15	.18	.32	.97	.94	.95
Y-76 ¹	.12	.18	.20
Y-84 ¹	.19	.20	.27
Log scale:						
Y-70-L ²	.20	.28	.37	.91	.85	.86
Y-71-L ²	.23	.27	.38	.92	.91	.90
Y-70-L ¹	.18	.29	.36
Y-71-L ¹	.24	.29	.31
Y-76-L ¹	.11	.15	.18
Y-84-L ¹	.14	.18	.23

¹ Correlations restricted to equal zero (independent analyses for each level).

² Values just prior to divergence.

a herd received more diverse diets now than in the past. These factors may be responsible for changes in variance components either from one level to another in a year or from one year to another.

Estimates of sire and residual components of variance for each year are plotted in Figures 2 and 3 as solid lines for linear scale analyses. Sire and residual components from the log scale analyses also are plotted in Figures 2 and 3 but as interrupted lines. Sire components from

the intermediate yr 1976 are smaller than the others. The estimates were generally similar for all three levels, but because residual components decreased in the higher levels, heritability was larger in the higher levels.

Residual components from log scale analyses did not change greatly over time. However, estimates for medium and high levels in 1984 are larger than those for other years when the means were also larger, especially for low and middle levels.

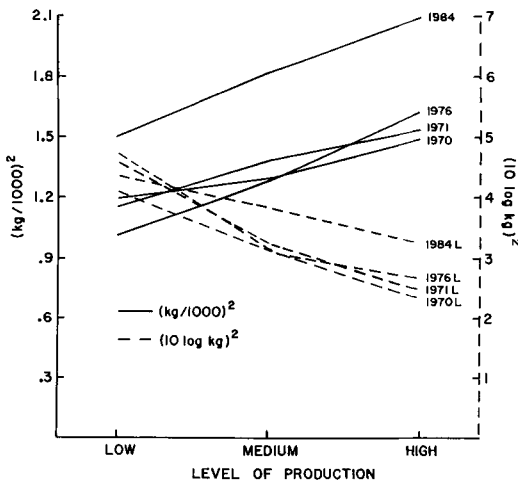


Figure 3. Estimates of residual variance components by level of production for each year on linear and logarithmic scales.

CONCLUSIONS

Sire and residual variance components increase importantly as production increases.

If records are transformed to logarithms, the sire component is essentially constant, but residual components of variance decrease as production increases.

Correlations between sire values at different levels seem for practical purposes to be near unity (all .85 or greater), which would confirm previous evidence indicating ranking of sires is not greatly affected by production level where daughters make their records.

Heritabilities increase as production increases, and consequently, selection would be more efficient when information comes from herds with higher average production.

Heritabilities for log transformed records are larger than for untransformed records, es-

pecially at higher production and earlier years. However, this result is not interpretable in terms of actual production, because components of variance, especially environmental ones, decreased as the level increased, seemingly due to inadequacy of multiplicative model to describe yields.

Effect on ranking of animals due to a model that does not account for both heterogeneity of variance components and heritability should be examined both for sire evaluation and for cow evaluation. Hill (10) has discussed the case of heterogeneous variances.

Powell et al. (19) concluded that genetic evaluations of cows (cow indexes) are insensitive to varying heritability and variances. Lofgren et al. (15) studied adjustments to cow indexes for herd mean and herd standard deviation from actual data and found the best adjustment was for both heritability and standard deviation. Since this paper was submitted, Gianola (13) has reviewed procedures for mixed model evaluation with heterogeneous genetic and residual variances. Similarly, Garrick and Van Vleck (5) also have outlined multiple trait evaluation procedures with environments considered as traits and have demonstrated the consequences of incorrectly assuming homogeneity using the results of Henderson (7) including the case of a singular genetic (co)variance matrix.

ACKNOWLEDGMENTS

This research was supported in part by grants from Eastern AI Cooperative, Inc., Ithaca, NY, and US-Spain Joint Committee for Scientific and Technological Cooperation. The senior author was supported by a scholarship from ODEPLAN and by Universidad Austral de Chile.

REFERENCES

- 1 Brotherstone, S., and W. G. Hill. 1986. Heterogeneity of variance amongst herds for milk production. *Anim. Prod.* 42:297.
- 2 Clay, J. S., W. E. Vinson, and J. M. White. 1979. Heterogeneity of daughter variances of sires for milk yield. *J. Dairy Sci.* 62:985.
- 3 Danell, B. 1982. Interaction between genotype and environment in sire evaluation for milk production. *Acta Agric. Scand.* 32:33.
- 4 Everett, R. W., and J. F. Keown. 1984. Mixed model sire evaluation with dairy cattle—experience and genetic gain. *J. Anim. Sci.* 59:529.
- 5 Garrick, D. J., and L. D. Van Vleck. 1987. Aspects of selection for performance in several environments with heterogeneous variances. *J. Anim. Sci.* 65:409.
- 6 Gianola, D. 1986. On selection criteria and estimation of parameters when the variance is heterogeneous. *Theor. Appl. Gen.* 72:671.
- 7 Henderson, C. R. 1975. Comparison of alternative sire evaluation methods. *J. Anim. Sci.* 41:760.
- 8 Henderson, C. R. 1984. ANOVA, MIVQUE, REML, and ML algorithms for estimation of variances and covariances. Page 257 in *Statistics: an appraisal*. H. A. David and H. T. David, ed. Iowa State Univ. Press, Ames.
- 9 Henderson, C. R. 1984. Applications of linear models in animal breeding. Univ. Guelph, Ontario, Can.
- 10 Hill, W. G. 1984. On selection among groups with heterogeneous variance. *Anim. Prod.* 39:473.
- 11 Hill, W. G., M. R. Edwards, M.K.A. Ahmed, and R. Thompson. 1983. Heritability of milk yield and composition at different levels and variability of production. *Anim. Prod.* 36:59.
- 12 Hintz, R. L., R. W. Everett, and L. D. Van Vleck. 1978. Estimation of genetic trends from cow and sire evaluations. *J. Dairy Sci.* 61:607.
- 13 Hudson, G.S.F., R. L. Quaas, and L. D. Van Vleck. 1982. Computer algorithm for the recursive method of calculating large numerator relationship matrices. *J. Dairy Sci.* 26:2018.
- 14 Lofgren, D. L., W. E. Vinson, R. E. Pearson, and R. L. Powell. 1985. Heritability of milk yield at different levels of herd mean and standard deviation for production. *J. Dairy Sci.* 68:2737.
- 15 Lofgren, D. L., W. E. Vinson, R. E. Pearson, and R. L. Powell. 1985. Adjustments to cow indexes for milk yield for effects of herd mean and standard deviation. *J. Dairy Sci.* 68:3301.
- 16 Mirande, S. L. 1984. Trends in genetic and phenotypic variances in milk, fat and fat test production in Holstein cattle. M.S. thesis, Cornell Univ., Ithaca, NY.
- 17 Mirande, S. L., and L. D. Van Vleck. 1985. Trends in genetic and phenotypic variances for milk production. *J. Dairy Sci.* 68:2278.
- 18 Powell, R. L., and H. D. Norman. 1983. Heritabilities of milk and fat yields according to herd-average yield. *J. Dairy Sci.* 66(Suppl. 1):123. (Abstr.)
- 19 Powell, R. L., H. D. Norman, and B. T. Weinland. 1983. Cow evaluation at different milk yields of herds. *J. Dairy Sci.* 66:148.
- 20 Robertson, A., K. O'Connor, and J. Edwards. 1960. Progeny testing dairy bulls at different management levels. *Anim. Prod.* 2:141.
- 21 Schaeffer, L. R., J. W. Wilton, and R. Thompson. 1978. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics.* 34:199.
- 22 Syrstad, O. 1976. Interaction between genotype and nutrition in dairy production. *World Rev. Anim. Prod.* 12:33.
- 23 Van Vleck, L. D. 1963. Genotype and environment in sire evaluation. *J. Dairy Sci.* 46:983.