PLISKA
STUDIA MATHEMATICA BULGARICA

ПЛИСКА
БЪЛГАРСКИ МАТЕМАТИЧЕСКИ СТУДИИ

For further information on
Pliska Studia Mathematica Bulgarica
visit the website of the journal http://www.math.bas.bg/~pliska/
or contact: Editorial Office
Pliska Studia Mathematica Bulgarica
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: pliska@math.bas.bg

Pliska Stud. Math. Bulgar. **17** (2005), 313–322

PLISKA
STUDIA MATHEMATICA
BULGARICA

COMPARISON OF MULTIVARIATE AND UNIVARIATE MODELS FOR GENETIC EVALUATION OF MILK YIELD BASED ON TEST DAY DATA

Yanka Tsvetanova

Multivariate and univariate lactation models were applied to test day data to predict genetic value of daily milk yield of a sample of Black and White cows. The models for genetic evaluation include a set of fixed main effects, fixed regression on functions of days im milk, random effects of permanent environment within lactation, random additive genetic effect and residual effect. Under multivariate model for daily milk yield test day records within lactation are considered as repeated measurements, and different lactations are treated as separate traits. Univariate model is applied for each lactation using test day yield as repeated measure. The variance components, genetic parameters and ranging of the animals through the multivariate and univariate metod were compared.

1. Introduction

Linear statistical models for genetic evaluation are widely applied in breeding programs for genetic improvement. The genetic evaluation of dairy sires and cows for productive traits has for many years been based on the analysis of 305-day lactation yields. The basis of every 305-day yield is a set of test day yields taken approximately every 30 days in milk. Cows may have from 2 to 12 test day measurements to provide a 305-day lactation measure. Incomplete lactation records are normally extended to 305-day basis following well defined rules. One

---

2000 *Mathematics Subject Classification:* 62H12, 62P99

Key words: mixed linear models, repeated measures, fixed regression, genetic parameters

way to avoid the problem of extension of test day yields to 305-day records is using test day yields for genetic evaluation of dairy sires and cows rather than 305-day yields. There are many advantages of test day models over the 305-day models. Ptak and Shaeffer [6] and Meyer et al. [4] demonstrated the advantages of the test day models over models based on 305-day measurement for production traits. Benefits are mainly the result of reduced residual variance through consideration of the effect of test day, which account for short - term environmental variation, instead oh herd-year-season classification, which assumes a common effect of herd and time on all test day observations. With test day models the accuracy of the genetic evaluation is better than with 305-day models.

The linear model assumed to explain test day milk yield would need to incorporate the general shape of the lactation curve, the different variation of test day yields depending on the number of the days in milk, the effects specific to a particular cow on a given test day, such as pregnancy or disease. A multivariate analysis involves the simultaneous evaluation of animals for daily milk production and takes into account the phenotypic and genetic correlations between the milk yield in adjacent lactations. One of the main advantages of multivariate animal model over the univatiate model for each lactation is increased accuracy of evaluations. The gain in accuracy depends on the absolute difference between the genetic and residual correlations between sequent lactations. The larger the difference in these correlations, the greater the gain in accuracy of evaluations [9]. Also there is an additional increase in accuracy resulting from better connection in the data, due to the residual covariances between milk yield in adjacent lactations [13].

During recent years linear models for dairy productive traits analysis based on test day records have been used in many investigations [12] and have been implemented in several developed countries as the USA, Germany, Canada, Australia, New Zealand for breeding value evaluation. Most of the researches on test day models have been carried out in countries with well-established breeding programs, official milk recording schemes and an accurate pedigree information. In our country these models are not so popular for some reasons. The total number of dairy cattle under observation in the recent years is relatively small. A national breeding program has not yet been established successfully. Official milk recording schemes have only been implemented in a small proportion of the cattle populations and test day data are not always available.

The presented study is an attempt to apply a mixed linear test day model with fixed regression to daily milk yield for genetic evaluation of cows and sires from Black and White breed. First we applied an univariate model with repeated test day measurments for each lactation and then a multivariate model where test day yields of different lactations were considered as different variables. The objectives of this study were: to compare variance components, genetic parameters, breeding value estimations for multiple lactations and single lactation model; to receive estimations of the regression coefficients of lactation curves for the different age-season of calving groups for the first three lactations.

2. Data and models

Data of Bulgarian Black and White cows from two herds located in the village of Medven in Sofia region and in the town of Chirpan collected from 1992 to 2000 were used in the study. 9930 test day records of milk yield for the first three lactations of 424 cows were included in the analysis. Data of the origin (dam and sire) and of age at calving, days in milk for each test day were used.

For genetic evaluation of daily milk yield records within a lactation were considered as repeated observations of the same variable and records from different lactations as observations of separate variables. In the mixed model for analysis of daily milk yield as random factors were included animal additive genetic effect and permanent environment effects by lactation. Fixed effects included in the model were herd-month-year of test date and a set of four covariables as functions of days in lactation estimated by parity, age and season of calving which accounted for the shape of the lactation curve. The estimate of permanent environmental effect for an animal represents environmental influences and non-additive genetic effects and they are peculiar to the animal and could have either positive or negative effect on its performance in life. Permanent environmental effects can be estimated only for animals with performance records. The sum of permanent environmental effect and additive genetic effect for each animal is termed the probable producing ability and present an estimate of the future performance of the animal in the same management conditions. This could assist farmers in selecting animals for future performance in the same herd [5]. For the univariate model with repeated measurements it is assumed that there is an additional relationship between records of an animal due to environmental factors or conditions that affect them permanently within lactation. For the multivariate model including first three lactations for analysis of test day records it is necessary to be considered both environmental effects of yields within lactation and permanent environmental effects between lactations.

In matrix notation the multivariate model can be written as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z_1}\mathbf{a} + \mathbf{Z_2}\mathbf{p} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{b}$ is the vector of fixed effects, $\mathbf{a}$ is the vector of random additive genetic effects, $\mathbf{p}$ is the vector of random permanent environmental effects and non-additive genetic effects for test day observation on cows within lactations, $\mathbf{e}$ is the vector of random residual effects, and $\mathbf{X}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$ are incidence matrices relating records to fixed, animal and permanent environmental effects respectively.

Expectations of random vectors are: $\mathbf{E(y) = Xb}$, $\mathbf{E(a) = E(p) = E(e) = 0}$. The assumptions under the model are that the permanent environmental effects and residual effects are not correlated. Covariance matrices of random effectss are: $\mathbf{V(a) = G = A \otimes G_0}$, $\mathbf{V(p) = P = I \otimes PE_0}$, $\mathbf{V(e) = R = I \otimes R_0}$, where $\mathbf{A}$ is the numerator relationship matrix between animals, $\mathbf{I}$ is the identity matrix, $\mathbf{G_0}$, $\mathbf{PE_0}$ and $\mathbf{R_0}$ denote 3x3 covariance matrices for additive genetic, permanent environment, and residual effects respectively.

Genetic evaluation under this model is concerned with predicting not only breeding values (additive genetic effects) but also permanent environmental effects. The BLUE (Best Linear Unbiased Estimation) of fixed effects ($\mathbf{b}$) and BLUP (Best Linear Unbiased Prediction) of random additive genetic effects ($\mathbf{a}$) and permanent environmental effects ($\mathbf{p}$) are solutions of Mixed Model Equations (MME), presented by Henderson [2]. For the multivariate model the matrix form of MME is:

$$\begin{pmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z_1} & \mathbf{X'R^{-1}Z_2} \\ \mathbf{Z_1'R^{-1}X} & \mathbf{Z_1'R^{-1}Z_1 \otimes G_0^{-1}} & \mathbf{Z_1'R^{-1}Z_2} \\ \mathbf{Z_2'R^{-1}X} & \mathbf{Z_2'R^{-1}Z_1} & \mathbf{Z_2'R^{-1}Z_2 + I \otimes PE_0^{-1}} \end{pmatrix} \begin{pmatrix} \mathbf{b} \\ \mathbf{a} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z_1'R^{-1}y} \\ \mathbf{Z_2'R^{-1}y} \end{pmatrix}$$

The estimations of fixed effects and predictions of the random effects in MME can be obtained when variance components are known. The most usual situation is when the variance components are unknown and they could be estimated by the various procedures for instance REML.

The multivariate linear model we applied is presented by the following equation:

$$Y_{ijkm} = HYM_{im} + a_{jm} + p_{jm} + AS_{km} + b_{km1}(D/c) + b_{km2}(D/c)^2 + \\ + b_{km3}ln(c/D) + b_{km4}ln(c/D)^2 + e_{ijkm}$$

where Y_{ijkmn} is the $i-th$ test day observation (milk in liters) of cow j in the $k-th$ age-season subclass of lactation m; HYM_{im} is the fixed herd-time effect (in our model herd-year-month of test day); a_{jm} is the random animal additive genetic effect for the trait in lactation m; p_{jm} random permanent environment effect associated with all test day records of cow j within lactation m; AS_{km} is the $k-th$ age-season subclass fixed effect in lactation m; b_{km1} and b_{km2} are fixed regression coefficients on linear and quadratic effects of D/c where D denotes days in milk on the test date and $c = 305$ (standard length of lactation), the second couple b_{km3} and b_{km4} are regression coefficients on linear and quadratic effects on $ln(c/D)$. Regressions are nested in age-season classes; e_{ijkm} is a random residual effect. For the univariate model the index m in the above equation is omited.

Univariate repeatability model for estimation of breeding value for milk yield was performed for every one of the first three lactations. Each univariate model provide separate genetic value estimation of daily milk yield for the respective lactation. With this model it is usually assumed that there is a genetic correlation of unity between all pairs of records of the same animal during lactation, that all records have equal variance and that the environmental correlations between all pairs of records are equal.

A drawback of classification by herd-testdate in a model is that the number of observations on a particular herd test date is very small for our data especially for second and the third lactation. This was the reason we use herd-month-year effect in the model. Herd-year-month effect partially accounts for short term systematic effects on environment on daily milk.

Regression coefficients on functions of days in milk were estimated within 24 groups by lactation, age of calving, and the season (three lactations, four groups for age of calving, and two seasons of calving. Similar model is proposed for the first time by Ptak and Shaeffer [6] for modeling productive traits in dairy cattle and later its variants have been applied by some other authors (Swalve [11], Reents et al. [7, 8]).

Variance components were initially estimated by REML procedure on the basis on analytical gradients implemented in the program VCE5 [3]. The received variance component estimations were used for computing BLUE of the fixed effects, coefficients of the fixed regression and BLUP of animal additive genetic effect and permanent environmental effect by the program PEST [1].

3. Results and discussions

Estimates of variance components and heritability for the univariate repeatability model for each lactation are shown in Table 1.

Table 1. Estimates of the variances of random effects and estimates of heritability of daily milk yield for 1,2,3 lactation.

Variances	Lactations		
	I	II	III
	n=4945	n=2979	n=1500
σ_a^2	4.39	5.71	4.48
σ_{pe}^2	1.09	1.49	1.93
σ_e^2	12.34	17.13	19.23
h^2	0.24	0.23	0.17
PE	0.06	0.05	0.07

σ_a^2 – animal additive genetic variance; σ_{pe}^2 – variance of permanent environment; σ_e^2 – residual variance; h^2 – heritability; PE – relative proportion of permanent environment to total variance; n – number of test day records.

Multivariate analysis requires reliable estimates of variance components for each lactation and covariance components between lactations. Their estimations are shown in Table 2.

Table 2. Estimates of variances and covariances of random effects of daily milk yield for lactations 1 to 3.

	Animal additive genetic			Permanent environment			Residual		
Lact.	I	II	III	I	II	III	I	II	III
I	5.07	4.82	4.61	4.95	2.50	1.45	11.31	5.56	6.86
II		6.28	5.66		5.50	2.03		16.62	10.31
III			5.51			3.61			19.71

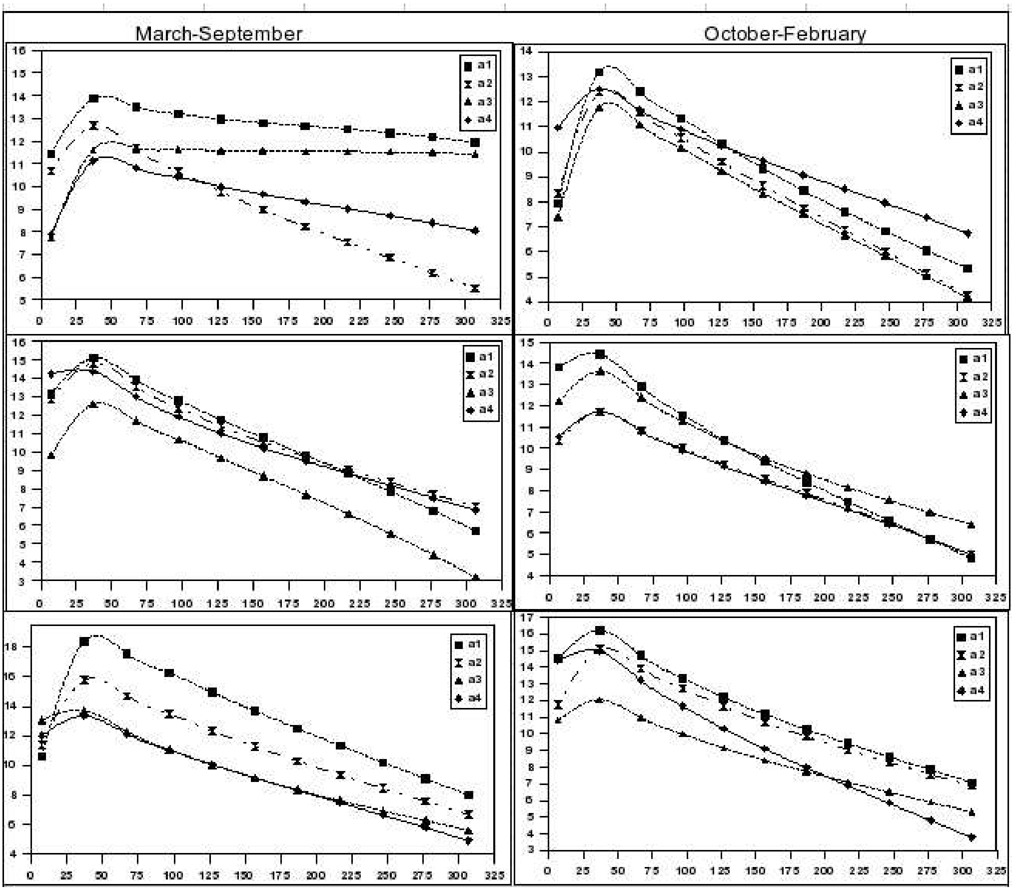
Genetic parameters of interest such are heritability , correlations between additive genetic effects and correlations betwenn permanent environment effects for different lactations. Genetic, permanent environment and residual correlations between lactations are shown in Table 3.

Table 3. Estimates of genetic correlations and heritability (on diagonal); correlations between permanent environmental effects; correlation between residual effects.

	Animal additive genetic			Permanent environment			Residual		
Lact.	I	II	III	I	II	III	I	II	III
I	0.24	0.85	0.81	0.23	0.48	0.42	0.53	0.41	0.31
II		0.22	0.88		0.18	0.46		0.60	0.43
III			0.21			0.12			0.66

With the multivariate model additive genetic variances are higher than those in the univariate repeatability model. The values of the heritability are slightly higher with the multivariate model. Heritability estimates for milk production decreased as lactation number increased. Genetic correlations between lactations are high. These results are in agreement with the investigations of other authors [10].

Figure 1: Graphics of the estimated lactation curves for the 1,2,3 lactations, grouped by age and season of calving.



In this study fixed regressions were nested within parity, age, and season of calving. For each lactation there are four groups for age of calving, noted by $a1$, $a2$, $a3$ and $a4$. They embrace the following age intervals in months: for the first lactation $a1 = [18 \div 23]$, $a2 = [24 \div 29]$, $a3 = [30 \div 35]$, $a4 = [36 \div 48]$; for the second lactation $a1 = [30 \div 35]$, $a2 = [36 \div 40]$, $a3 = [41 \div 45]$, $a4 = [46 \div 53]$; for the third lactation $a1 = [40 \div 46]$, $a2 = [47 \div 53]$, $a3 = [54 \div 60]$, $a4 = [61 \div 68]$. Figure 1 shows fitted lactation curves for milk production. Curves were adjusted to a fixed effects and do not display the increase in absolute level of milk. Curves for the season "March-September" in first parity differed significantly from curves for second and third parities. Across age and season group curves differed mainly in heigh but were similar in shape. Slight seasonal effect in regression coefficient is noted and this result in slight higher productivity in the season "March-September" than in the season "October-February". On the other hand the persistence of the lactation curve is more stable for spring-summer season for the first lactation while for the next lactations the shape of the lactation curves are similar.

Table 5. Example listing of estimated breeding values for milk yield of sires with more than 50 daughters obtained by multivariate model and univariate model for the separate lactations and rank correlations.

	Lactation I		Lactation II		Lactation III	
	$R = 0.97$		$R = 0.96$		$R = 0.94$	
Animal	MVM	UVM	MVM	UVM	MVM	UVM
30292	1.2599	1.5611	−3.2804	−3.6661	1.4308	1.9272
32592	−0.572	−0.3844	2.0315	2.1853	−2.3659	−1.7678
33192	−1.479	−1.0592	−0.5925	−0.2248	3.6320	4.6402
33792	0.2301	0.3918	−2.203	−2.8103	0.0831	−0.8855
35492	−2.924	−1.6139	−0.2811	−0.3188	1.0203	1.5139
35892	−1.1151	−0.4146	2.3802	1.8507	−0.3715	−0.8389
36392	−1.9452	−1.5889	−1.1725	−1.6108	0.0693	−0.2313
38192	0.6072	1.1138	−1.6596	−2.4536	−2.3573	−2.3944
38292	1.8316	1.0898	−5.5057	−6.6600	−3.7104	−5.0725
38792	0.1847	0.3945	1.0056	0.5683	0.5450	0.4787
40292	1.0498	1.2624	1.8935	2.3804	−0.5885	−0.3750
40692	−1.7154	−0.3294	−0.1899	−0.2149	1.023	0.5107
40892	−0.7449	−0.252	2.6015	3.3073	3.4315	3.7387
41592	0.6474	0.7768	1.0233	1.4647	1.5908	1.3570
41992	1.4491	1.1489	−2.4461	−2.0238	0.6364	0.6041

For production traits most attention is paid to the ranking of sires with the highest breeding value estimations. Sires may be more accurately proven by having a large number of test day yields available on their daughters. Differences between estimated breeding values for milk production obtained by multivariate model and univariate model for separate lactations were small for bulls with more than 50 daughters. The Spearman correlation coefficients between bulls ranks obtained by univariate and multivariate method are high. They are shown in Table 5.

4. Conclusions

The multiple-lactation model with repeated observations within a lactation with fixed regression may not entirely account for the fact that adjacent test days are more highly correlated than distant test days. Methods to account more properly for heterogeneous correlations between test day observations include the use of random lactation curves. That model will be the object of our further studies.

REFERENCES

[1] Groeneveld E. PEST User's Manual, Institute of Animal Husbandry and Animal Science, Mariensee, Germany, 1990.

[2] Henderson C.R. Application of linear models in animal breeding. University of Guelph, 1984.

[3] Kovach M., E. Groeneveld VCE-5. User's Guide and Reference Manual. Ver. 5.1, Institute of Animal Husbandry and Animal Science, Mariensee, Germany, 2002.

[4] Meyer K., H.U. Graser, K. Hammond Estimates of genetic parameters for first lactation test day production of australian Black and White cows *Livest. Prod. Sci.* (1989), 21:177–199.

[5] Mrode R.A. Linear models for the prediction of animal breeding values, CAB International, Wallingford, UK, 1989, 67–73.

[6] Ptak E., L.R. Shaeffer Use of test day for genetic evaluation of dairy sires and cows. *Livestock Prod. Science* (1993), 34:23–34.

[7] Reents R., J.C.M. Dekkers, L.R.Shaeffer Genetic Evaluation for Somatic Cell Score with a test day model for Multiple Lactations. *J.Dayry Sci.* (1995), 78:2858–2870.

[8] Reents R., J. Jamrozik, L.R.Shaeffer, J.C.M. Dekkers Estimation of genetic parameters for test day records of somatic cell score. *J.Dairy Sci.* (1995), 78:2847–2857.

[9] Shaeffer L.R. Sire and cow evaluation under multiple trait models. *J.Dairy Sci.* (1984), 67:1567–1580.

[10] Stanton T. L.,L.R. Jones, R.W. Everett,S.D. Kachman Estimating milk, fat , and protein lactatoion curves with a test day model. *J.Dairy Sci.* (1992), 75:1691–1700.

[11] Swalve H. H. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* (1995), 78:929–938.

[12] Swalve H. H. Theoretical basis and computational methods for different test day genetic evaluation methods. *J. Dairy Sci.* (2000), 83: 1115–1124.

[13] Thompson R. , K. Meyer A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Prod. Science* (1986), 15:299–313.

Yanka G. Tsvetanova
Trakia University,
Faculty of Agriculture,
Dept. of Informatics, Mathematics and
Physics,
6000 Stara Zagora,
e-mail: `yanka@uni-sz.bg`