Trakia Journal of Sciences, Vol. 2, No. 3, pp 65-69, 2004

Available online at:
http://www.uni-sz.bg
ISSN 1312-1723

Original Contribution

LINEAR MODELS FOR BREEDING VALUE ESTIMATION OF DAIRY CATTLE BASED ON TEST DAY RECORDS

Yanka Tsvetanova*

Department of Informatics, Mathematics and Physics, Faculty of Agriculture,
Trakia University, Stara Zagora, Bulgaria

ABSTRACT

Linear statistical models for breeding value estimation are widely applied in breeding programs for genetic improvement. The evolution of computer technologies and computational power increase allow realization of the advanced methods for breeding value evaluation. During recent years linear models for dairy productive traits analysis based on test day records have been used in many investigations and have been implemented in several developed countries as USA, Germany, Canada, Australia, New Zealand for breeding value evaluation.

The aim of the current study is to present the theoretical basis of the models for breeding value estimation and their application of test day records for productive traits, genetic origin and the reproductive traits of Bulgarian Black and White cows.

Key words: repeatability test day models, variance components, Black and White cattle

INTRODUCTION

In general, Animal linear models for 305d lactation have been used for analysis of the productive traits in dairy cattle. During the recent years, in the place of the traditional complete lactation models, Test-day models gained a considerable usage for breeding value estimation of traits related to milk production. Test day models have been defined as a statistical procedure that considers genetic and environmental effects on a test day basis (1). There are several advantages of the Test day models over the traditional 305d lactation models. Among these advantages are the ability to account for environmental effects of each test day, the ability to model the trajectory of lactation for individual genotypes or groups of animals, avoidance of the use of extended records for culled cows and for records in progress. The disadvantages of the Test day models are: the amount of analyzed data is much larger and, secondly, many parameters have to be estimated compared with the models for complete lactation yields (2).

*Correspondence to: Yanka Tsvetanova, Department of Informatics, Mathematics and Physics, Faculty of Agriculture, Trakia University, Stara Zagora

Test day models are based on consecutive measures on the same trait once every 30 days in lactation on each animal. 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. Various statistical models have been proposed for analysis of test day records and they include: a repeatability models (1), multiple trait models (3) and random regression models (4).

Repeatability linear models are presented by the following matrix equation:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{pe} + \mathbf{e} \quad [1]$$

where

- **y** is a vector of observations;
- **β** is a vector of fixed effects;
- **a** is a vector of random additive genetic effects;
- **pe** is a vector of random permanent environmental effects and non-additive genetic effects;
- **e** is a vector of random residual effects;
- **X**, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are incidence matrices relating observations to fixed, animal and permanent environmental effects respectively.

The assumptions under the model are that the permanent environmental effects and residual effects are not correlated, with means of zero and variances σ_{pe}^2 and σ_e^2 respectively. The variance covariance matrices of the random effects are

$Var(pe)=I\sigma_{pe}^2$; $Var(e)=I\sigma_e^2$; $Var(a)=A\sigma_a^2$; $Var(y)=Z_1AZ_1'\sigma_a^2 + Z_2I\sigma_{pe}^2 Z_2' + R$.

A is the matrix of the additive genetic relationship among animals, called the numerator relationship matrix.

With this model it is usually assumed that there is a genetic correlation of unity between all pairs of records of the same animal, that all records have equal variance and that the environmental correlations between all pairs of records are equal. The variance of the observations comprises the genetic variance, variance due to permanent environmental effect and variance due to random temporary environmental effects ($\sigma_y^2 = \sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2$). The correlation between records of an animal, referred as a repeatability, is $\mathbf{r} = (\sigma_a^2 + \sigma_{pe}^2)/\sigma_y^2$.

Compared with the 305d lactation models repeatability Test day models produce more accurate prediction of breeding value. The gain in accuracy results mainly from the reduction of temporary environmental variance as the numbers of the records of an animal on the same trait are more than one. When repeatability is low this gain is considerable as number of records increase. When repeatability is high, the gain in accuracy is not substantial from repeated record models compared with selection on single record (5). Genetic evaluation under this model is concerned with predicting not only breeding values but also permanent environmental effects. The BLUE (Best Linear Unbiased Estimation) of fixed effects and BLUP (Best Linear Unbiased Prediction) of random additive genetic and permanent environmental effects are solutions of Mixed Model Equations (MME), presented by Henderson (6):

$$\begin{pmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1'Z_1+\lambda_A A^{-1} & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2+\lambda_p I \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{a} \\ \hat{p} \end{pmatrix} = \begin{pmatrix} X'y \\ Z_1'y \\ Z_2'y \end{pmatrix} \qquad [2]$$

where $\lambda_A = \sigma_e^2 / \sigma_a^2$; $\lambda_p = \sigma_e^2 / \sigma_{pe}^2$.

The estimations of fixed effect 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. VCE5 (7) is one of the programs for dispersion components estimation and PEST (8) is one of the programs for estimation of the fixed effects and prediction of random effects under the linear models.

Selection in dairy cattle is usually based on a combination of several traits of economic importance which may be phenotypically and genetically related. A multiple trait analysis involves the simultaneous evaluation of animals for two or more traits and takes into account the phenotypic and genetic correlations between the traits. One of the main advantages of multivariate animal model is that it increases the accuracy of evaluations. The gain in accuracy is dependent on the absolute difference between the genetic and residual correlations between the traits. The larger the difference in these correlations, the greater is the gain in accuracy of evaluations (4).

Most of the research on Test day methods has been carried out in countries with well-established breeding programs, official milk recording schemes, accurate pedigree information. In Bulgaria these models are not so popular for some reasons. The total number of dairy cattle 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. Pedigree information and test day data are not always available.

The objective of this analysis is an attempt to determine genetic and environmental factors affecting daily milk yield and fat percent in milk for a sample of Black and White cattle using test day repeatability models for the first three lactations.

MATERIAL AND METHODS

Data of Bulgarian Black and White cows from two herds in the village of Medven in Sofia region and in the town of Chirpan, Bulgaria were used in the study. 9930 test day records for productive traits milk yield and fat percentage in milk from the first three lactations of 424 cows were included in the analysis. Data of the origin (dam and sire) and of reproductive traits as age at calving, days in milk for each test day, period of lactation, period of pregnancy on the test day were used.

In our study we used repeatability model as the most widely used models for the analysis of test day records and which are not so demanding of computational power.

The choice of fixed environmental effects that affect test day production is an important moment in the development of breeding value evaluation models. In general, the fixed effects that affect test day production are not quite different from those that influence the aggregated dairy records of complete lactation. The HTD (herd-test-day) is commonly used to account for the effect of the herd and the year of production and covers the season of production. Another effects usually included in the test day models are age at calving and the effect of the lactation. In the course of lactation, pregnancy has a significant impact on dairy production (9).

To find the fixed effects that affect significantly the milk performance data we used ANOVA (Analysis of Variances – Main effects ANOVA) procedure implemented in the statistical package StatSoft Statistica 6.0. Comparing various models with combinations of fixed factors that are likely to influence milk yield and fat percent in milk on the test day we found model (I) as the most appropriate for the data we used.

$$y_{ijklmn} = HYM_i + LAKT_j + AGE_k + PP_l + PL_m + e_{ijklmn} \qquad (I)$$

- $\mathbf{y_{ijklm}}$ are the observation of the dependent trait (test day milk yield or fat percent in milk);
- $\mathbf{HYM_i}$ is a fixed effect of the i-th level of the herd-year-month of the test day. контролата;
- $\mathbf{LAKT_j}$ is a fixed effect of the corresponding lactation;
- $\mathbf{AGE_k}$ is a fixed effect of the corresponding age group of first calving;
- $\mathbf{PP_l}$ is a fixed effect of the corresponding pregnancy period;
- $\mathbf{PL_m}$ is the fixed effect of the corresponding period of lactation;
- $\mathbf{e_{ijklmn}}$ is the random environmental effect.

RESULTS AND DISCUSSIONS

For model (I) were received the highest coefficient of determination R^2 and the lowest error variance. The results from the analysis of variances are shown on **Table 1**.

When HYM (herd-year-month) of the test day is included in the model (I) as a random factor, the variance components estimations are obtained by the REML procedure of Statistica 6.0. These variance estimations are shown on **Table 2**.

Table 1. The ANOVA results for determining the influence of fixed factors on the productive traits of test day for the first three lactations.

Source of variation	df	Milk yield model		Fat % in milk model	
		MS	F	MS	F
Age of first calving	7	95.2	8.79***	0.8	2.3*
Lactation	2	1204.2	111.25***	0.5	1.4
Herd-year-month	114	271.6	25.09***	8.0	22.3***
Period of lactation	9	2536.5	234.33***	21.7	60.5***
Period of pregnancy	8	466.7	43.12***	2.6	7.3***
Error	8816	10.8		0.4	
		$R^2 = 0.46$		$R^2 = 0.50$	

*- p<0.05, ** - p<0.01, *** - p<0.001

Table 2. Variance components of test day model (I) with HYM random factor for milk yield and fat % in milk.

	Milk yield model			*Fat % in milk model*		
	σ^2	*Std.Err*	*p-value*	σ^2	*Std.Err*	*p-value*
Herd-year-month variance	3.85	0.54	*<0.001*	0.11	0.015	*<0.001*
Error variance	10.82	0.16	*<0.001*	0.36	0.005	*<0.001*
Phenotype variance	*14.67*			*0.47*		

Comparing the results on **Table 1** and **Table 2** we found that when the factor Herd-year-month of the test day was considered as a random in the models for milk yield and for fat percent in milk the value of the error variance was almost the same as in the models with fixed HYM.

The test day records of each animal are

subsequent records on the same trait and should be considered as repeated records. At the next step, for modeling test day milk yield and fat percentage in milk we applied animal repeatability model (II).

$$y_{ijklmn} = HYM_i + LAKT_j + AGE_k + PP_l + PL_m + a_n + pe_n + e_{ijklmn} \quad (II)$$

In the model (II) are included the same fixed factors as in the model (I) (HYM, LAKT, AGE, PP, PL). In the model (II) a_n is the random animal additive genetic effect and pe_n is the random permanent environmental effect of the same animal and e_{ijklmn} is the random temporary environmental effect.

Animal model allows us to estimate simultaneously the fixed effects and a breeding value for each animal in the population. For estimation of the animal breeding value data for genetic relationship between animals were used. Breeding value estimation for the productive traits of each animal is based on its own records as well as on records of its relatives. Estimates of breeding values can be used to rank animals for genetic purposes.

Permanent environmental effects **(pe)** are usually accounted for in the repeatability models to ensure accurate prediction of breeding value. The estimate of environmental effect for an animal represents environmental influences and a non-additive genetic affects, and it is specific for an animal and affects its performance for life. The differences in estimates of **pe** represent permanent environmental and non-additive genetic differences between animals and could assist the farmer, in addition to the breeding value, in selecting animals for future performance in the same herd (5). The sum of breeding value estimate and the estimate of permanent environmental effect of the same cow is termed the probable producing ability (PPA) and can be used to estimate the future performance of the cow, which would be useful for the purposes of culling.

The estimations of random effects variances of the model (II) are shown on **Table 3**. They are obtained by the program for variance component estimation VCE 5 (7). Comparing with the results on **Table 2** we observed that the inclusion of the additive genetic effect and permanent environmental effect in the model for the productive traits decreased the error variance estimation. Estimates of the additive genetic variance and phenotypic variance can be used for estimation of the heritability (h^2). Heritability is the proportion of the phenotypic variance that is due to additive genetic effects. From **Table 3** the estimates of the heritability for the milk yield h^2=0.26 and for the fat % in milk h^2=0.52 could be done. The repeatability for the milk r = 0,52 and for fat percent r= 0.76. The estimations of the fixed factors and predictions of breeding values and permanent environmental effect are obtained by the program PEST (8).

Table 3. Variance components of repeatability test day animal model (II) for milk yield and fat % in milk.

	Milk yield model		*Fat % in milk model*	
	σ^2	*% of the phenotype variance*	σ^2	*% of the phenotype variance*
Additive genetic variance	2.96	26	0.073	52
Variance due to permanent environmental effects	3.13	27	0.033	24
Error variance	5.61	47	0.033	24
Phenotype variance	*11.7*		*0.139*	

Table 4. Variance components of multivariate model on milk yield and fat % in milk.

	Milk yield model		*Fat % in milk model*		
	σ^2	*% of the phenotype variance*	σ^2	*% of the phenotype variance*	*Correlations*
Additive genetic variance	2.93	25	0.084	57	-0.42
Variance due to permanent environmental effects	3.15	27	0.028	20	-0.39
Error variance	5.61	48	0.034	23	-0.18
Phenotype variance	*11.69*		*0.146*		

At the next step of our investigation we implemented multivariate analysis for the two traits – milk yield and fat percent in milk for test day using the same models for both traits as a univariate model (II), thus a multivariate model with equal design matrices. The variance component estimations for the multivariate model are shown on Table 4 as well as the corresponding genetic permanent environmental and error correlations between traits. There is a slight difference in variance estimations in comparison with the univariate models for the separate traits. We found that the differences in animal ranging based on univariate models and multivariate model are similar.

The corresponding heritability values for milk yield is h^2=0.25 and for the fat percent in milk is h^2=0.57. The repeatability for the milk r = 0,43 and for fat percent r= 0.76.

CONCLUSIONS:

The effect of herd- year-month of test day should be considered as fixed because when it is included as a random in the model the error variance does not change significantly.

The repeatability test day univariate model for milk yield and fat percent in milk increase the accuracy of estimations because it accounts for the permanent environmental factors that influence productive traits for each lactation.

Multivariate model on milk yield and fat percent in milk applied to data for Black and White cattle gives similar variance component values and similar ranging of the animals as in the separate univariate models for the two productive traits.

REFERENCES:

1. Ptak , E. , L.R. Shaeffer, Use of test day yields for genetic evaluation of dairy sires and cows, *Livestock Prod. Sci*. 34:23-34,1993.
2. Jensen J., Genetic evaluation of dairy cattle using Test–day models, *J. Dairy Sci.*, 84:2803-2812, 2001.
3. Reents, J.C.M. Dekkers, SL.R.Shaeffer, Genetic Evaluation for Somatic Cell Score with a test day model for Multiple Lactations, *J.Dayry Sci.,* 78:2858-2870, 1995.
4. Shaeffer L. R., Sire and cow evaluation under multiple trait models, *J. Dairy Sci*. 67: 1567-1580, 1984.
5. Mrode, R.A., Linear models for the prediction of animal breeding values, CAB International, Wallingford, UK, 1998.
6. Henderson, C. R., Application of Linear Models in Animal Breeding, Univ. of Guelph Press, Guelph, Canada, 1984.
7. 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.
8. Groeneveld E., PEST User's Manual, Institute of Animal Husbandry and Animal Science, Mariensee, Germany, 1990.
9. Swalve, H. H., Theoretical basis and computational methods for different test-day genetic evaluation methods. *J. Dairy Sci.* 83: 1115-1124, 2000.