Comparison of Different Lactation Curve Sub-Models in Test Day Models

Z. Guo and H.H. Swalve
Institut für Tierzucht und Haustiergenetik der Universität Göttingen
Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany

Abstract

The goodness of fit of 16 lactation curve functions was evaluated using field data of daily milk recordings collected on three farms and considering various criteria. The fit of lactation curve functions was improved for second and later lactations compared to first lactations. Differences between farms were substantially higher than those between lactation curves functions. For the estimation of variance components using regular milk recording data in a second part of the study 13 different functions were used as sub-models in test day models. Estimates of variance components were virtually identical for all functions. Furthermore, hardly any changes in rank from top lists of sires and cows occurred when comparing estimated breeding values under the models used.

Introduction

In test day repeatability models (Ptak and Schaeffer, 1993) the effect of the stage of lactation commonly is accounted for by a set of covariables that may be denoted as a sub-model. Sub-models can be taken from the numerous studies dealing with modeling the lactation curve. Guo and Swalve (1995) compared the goodness of fit of 14 models using data of daily milk recordings from an experimental farm and considering various criteria for comparison. Aim of the present study was to redo this comparison using field data from three large dairy farms in North-Eastern Germany and furthermore estimate variance components under test day models differing by the choice of the sub-model using conventional milk recording data.

Material and methods

In January of 1996 a project was initiated to collect daily recordings of milk yield from the on-farm computers of three dairy farms in North-Eastern Germany. The size of the herds was 1000, 450, and 3000 cows, respectively. In a first attempt to analyze the data, data collected up to June 1997 (farms A and B) and up to March 1997 (farm C) were used. As a consequence, complete lactations of cows were only available for a fraction of cows of each herd that fitted into the given time frame. More data will accumulate as the collection continues.

Data was edited with respect to the number of missing days per lactation (# 10 days in total), the number of consecutive days missing (# 6 days), the interval between date of calving and the start of the recording (# 9 days) and the length of lactation ($ 250 and # 450 days).

Table 1 displays the structure of the data. Average production (highest on farm B) and quality of the data varies from farm to farm. Missing days of recording exist due to misidentification of cows in the milking parlor and cows affected by diseases being taken off milking in the main parlor. Yield at missing days of recording was interpolated using the method of natural cubic splines to ensure that all lactations that went into further analysis were complete. The final data set was used to mimic a standard sampling method with a start at day 10 and regular intervals of 30 days for all cows.

Table 1. Structure of the data of daily milk recording collected on three farms.

Farm	Pa-		LL[2](days)				Milk Production (kg)				NM[3]		ICT[4]	
	rity	n[1]	0[5]	s.d.[6]	min	max	0	s.d.	min	max	0	s.d.	0	s.d.
A	1	116	302	35	260	431	4946	1184	2923	8123	6.0	1.8	5.7	1.1
	2	81	300	29	260	395	6262	1292	3063	9200	6.1	1.4	5.5	1.1
	3+	55	313	40	262	439	6892	1199	4839	9734	6.6	1.9	5.5	1.1
	T[7]	252	303	35	260	439	5794	1466	2923	9734	6.2	1.7	5.6	1.1
B	1	58	311	47	254	435	6467	1608	3861	11645	1.0	2.0	4.8	1.8
	2	36	349	50	290	450	8686	2045	4743	13768	1.4	2.2	3.9	1.6
	3+	73	329	48	254	450	7973	1874	4849	12850	0.8	1.5	4.7	1.5
	T	167	327	50	254	450	7604	2015	3861	13768	1.0	1.8	4.6	1.6
C	1	315	313	41	258	432	5161	1098	2603	9234	3.3	2.6	6.6	1.5
	2	126	301	33	258	413	5800	1124	2651	8530	3.7	2.2	5.7	1.3
	3+	150	307	41	257	427	6187	1382	3505	11699	3.6	2.3	6.3	1.5
	T	591	309	39	257	432	5558	1261	2603	11699	3.4	2.4	6.3	1.5

[1]Number of lactations [2]Length of Lactation [3]Number of missing days per lactation [5]Mean
[4]Interval between calving date and first test date [6]Standard Deviation [7]Total

A total of 16 lactation curve functions, 4, 5, and 7 functions with 3, 4, and 5 parameters, respectively, were fitted. Details and definition of functions can be found in the appendix. The goodness of fit was evaluated considering various criteria. Only results for the criteria **R** (correlation between estimated and true yield per test day), **AE** (absolute error computed from the difference of true and estimated yield per test day), and **M1** / **M** (difference between true and estimated values of cumulative yield of first 100 days / entire lactation) shall be reported here. Differences between functions with respect to the four criteria were analyzed separately for parity, farm, and criterion with an ANOVA applying a model that included FUNCTION (fix) and COW (random) as the only two independent variables. The Tukey-Test was chosen to test differences between models.

Of the 16 lactation curve functions, 13 functions were used as sub-models in a test day model of the following form:

$$y_{ijkl} = HTD_i + JSA_j + [f_j(b_{jm}, DIM_{kl})] + a_k + pe_k + e_{ijkl}$$

where

y_{ijkl} = test day milk yield
HTD_i = fixed effect of herd-test-day (HTD)
JSA_j = fixed effect of year-season-first calving age (180 levels)
a_k = cow effect(random)
pe_k = random permanent environmental effect
e_{ijkl} = residual effect
b_{jm} = regression coefficient
DIM_{kl} = days in milk
$f_j(b_{jm}, DIM_{kl})$ = function of DIM describing the shape of the year-season-first calving age subclasses lactation curve (sub-model)

The data used for the estimation of variance components consisted of 14,756 first lactations from a regular milk recording scheme, the data set being identical to the one used by Swalve (1995). REML estimation was carried out using VCE 3.2 (Groeneveld, 1996).

Results and discussion

A comparison of the results for the goodness of fit using the data collected on farm showed that differences between farms and parities within farms were larger than those found between the lactation curve functions. In general, a better fit was found for second and later parities than for first lactations. Table 2 displays the results for first parities, the functions being grouped by the number of their parameters (3, 4, 5, respectively):

Table 2. Comparison of lactation curve functions across farms (First lactations).

Function	Farm A				Farm B				Farm C			
	R	AE	M1	M	R	AE	M1	M	R	AE	M1	M
WOOD	0.719	1.934	42.9	56.5	0.549	2.156	48.7	72.0	0.676	2.231	58.4	69.9
W2	0.736	1.776	27.0	8.0	0.567	2.025	23.7	14.5	0.698	2.076	25.0	2.8
LM	0.743	1.757	13.5	-1.9	0.575	2.012	14.7	8.2	0.709	2.041	7.2	-11.1
MIL1	0.738	1.764	24.5	4.8	0.569	2.024	20.6	16.2	0.700	2.070	21.3	- 0.8
W1	0.724	1.768	31.3	3.6	0.580	2.006	-14.6	8.1	0.685	2.103	19.2	1.1
LM1	0.733	1.734	19.6	-5.8	0.585	2.006	-13.8	5.5	0.698	2.059	6.3	-11.9
LM2	0.728	1.747	23.7	-4.4	0.583	2.015	-19.4	2.3	0.692	2.069	5.9	-11.7
MIL4	0.730	1.751	25.5	-0.3	0.584	2.006	-10.0	10.0	0.692	2.086	15.6	- 3.0
MIL5	0.729	1.755	22.3	-0.0	0.586	1.998	-11.6	9.0	0.691	2.087	10.6	- 4.7
AS	0.703	1.916	15.9	-12.3	0.558	2.211	-16.7	- 1.8	0.673	2.262	- 7.3	-19.2
MW2	0.722	1.755	20.9	-5.4	0.572	2.046	-14.3	5.0	0.689	2.117	18.8	4.0
MK1	0.722	1.753	16.7	-9.6	0.578	2.014	-15.0	1.3	0.686	2.120	0.7	- 5.8
NEW5	0.722	1.749	17.9	-6.8	0.578	2.008	-15.2	3.0	0.686	2.130	4.6	- 0.6
NEW6	0.719	1.773	22.4	-1.8	0.570	2.044	-12.9	6.4	0.681	2.165	18.3	10.6
MIL6	0.724	1.745	16.4	-9.1	0.580	2.018	-15.4	3.2	0.690	2.114	7.1	- 2.0
MIL7	0.721	1.769	22.7	-1.9	0.569	2.053	-8.8	8.2	0.687	2.140	23.3	8.1
MSD[1)]	0.023	0.077	16.9	12.4	0.025	0.123	24.2	17.7	0.013	0.041	13.0	8.4

[1)]Minimum significant difference (Tukey-Test)

The effect of FUNCTION in the ANOVAs within farm was significant for all criteria and farms. However, as can be seen from Table 2, differences are marginal between models. An especially remarkable finding is the relatively low correlation between estimated and true yield for farm 2 which had the highest producing cows (see Table 1). The criteria R and AE refer to the capability of a model to estimate the daily yield at a specific test day whereas M1 and M consider the estimation of accumulated yield. For M1 and M over- and under-estimation of a specific lactation can cancel themselves out when averages are taken across lactations. This would suggest to use a criterion that considers absolute values of deviations. However, M1 and M are used here to have some insight in a systematic over- or underestimation. The use of a criterion AM considering the absolute deviation when estimating lactation yield (not shown in Table 2) gave values of between 160 to 180 thus suggesting that the error in estimating the lactation yield would be around 3%.

The results given in Table 2 show that even functions with only three parameters with the exception of Wood=s model can do well in fitting lactation curves. This would suggest to use simple functions as sub-models in test day models, a suggestion, that already has been subject to further analysis for random regression test day models (Jamrozik et al., 1997).

The results from the estimation of variance components for the official milk recording data applying 13 models differing by the choice of the sub-model can be summarized in a very brief way: Virtually all models lead to nearly identical estimates of the residual and the genetic component and thus to almost identical heritability estimates between 0.27 and 0.28. This leads to the conclusion that fixed regression coefficients to account for the curvilinear pattern of lactation yield do not affect the separation of environmental and genetic effects. Furthermore, from an inspection of the breeding values estimated under the 13 different models, hardly any changes in rank from model to model could be observed on top lists for cows and sires. Under random regression test day models (Schaeffer and Dekkers, 1994; Jamrozik and Schaeffer, 1997) some effect of the choice of sub-models may be observed (Jamrozik et al., 1997).

For the sake of comparison, a further test day model was used to estimate variance components and breeding values. This model was identical to the general one used before with the exception that the sub-model was dropped completely. The estimate of the residual variance increased by 15% compared to the 13 models used before. However, the additive genetic component also increased slightly resulting again in a heritability estimate of 0.27. Considering the estimated breeding values for cows and sires some but no dramatic changes in ranks compared to the models used before occurred.

Conclusion

Differences in the goodness of fit of lactation curve functions exist and are worthwhile to be exploited in a way that simple functions with few parameters that still have a sufficient fit may be used in test day models. From the field data collected on-farm it is clear that differences between farms (and presumably between individual cows) are of great importance. This leads to the conclusion that sub-models should be used as individual as possible, i.e. as suggested by Jamrozik and Schaeffer (1997) who in a first step apply the sub-model in a fixed way nested within contemporary groups that ideally should be herds or herd-years. This, however, may only be applicable to very large herds. The second step in an individual use of sub-models clearly is modeling individual curves in random regression models as suggested by the same authors. Comparisons of random regression models with ordinary test day models have been made, however, nothing has yet been published on differences in ranking of sires and cows under both approaches. The results in the present study suggest that changes in rank should occur due to large differences between individual farms and individual cows.

References

Groeneveld, E. 1996. VCE 3.2, ftp://ftp.tzv.fal.de/pub/vce3.2.

Guo, Z. and Swalve, H.H. 1995. Interbull Mtg. Prague, Uppsala, *Interbull Bull. No. 11*.

Jamrozik, J. and Schaeffer, L.R. 1997. *J. Dairy Sci. 80*, 762-770.

Jamrozik, J., Kistemaker, G.J., Dekkers, J.C.M. and Schaeffer, L.R. 1997. *J. Dairy Sci. 80*, (accepted).

Ptak, E. and Schaeffer, L.R. 1993. *Livest. Prod. Sci. 34*, 23-34.

Schaeffer, L.R. and Dekkers, J.C.M. 1994. *Proc. 5. WCGALP*, Guelph, *Vol. 18*, 443-446.

Swalve, H.H. 1995. *J. Dairy Sci. 78*, 929-938.

Acknowledgement

The authors gratefully acknowledge the participation of the three farms for the provision of their data and the support of the herdbook association of Mecklenburg-Vorpommern (RMV). Further gratitude is expressed to VIT, Verden, for additional supply of data.

Appendix

I. Models with 3 parameters

1. Wood (1967): $y_t = a_1 t^{a_2} e^{a_3 t}$ (**W**)

2. Wilmink II(1987): $y_t = a_1 + a_2 t + a_3 e^{-0.05t}$ (**W2**) ⊕

3. Logarithm-Model (Guo & Swalve, 1995): $y_t = a_1 + a_2 t + \frac{a_3}{t} e^{-\frac{1}{2}(\frac{\lg t - 1}{\sigma})^2}$ (σ=0.6) (**LM**) ⊕

4. Mixed-Log-Model I (Guo & Swalve, 1995): $y_t = a_1 + a_2 \sqrt{t} + a_3 \ln t$ (**MIL1**) ⊕

II. Models with 4 parameters

1. Wilmink I(1987): $y_t = a_1 + a_2 t + a_3 t^2 + a_4 e^{-0.05t}$ (**W1**) ⊕

2. Logarithm-Model I (Guo & Swalve, 1996):
   $y_t = a_1 + a_2 t + a_3 \sin(\frac{t}{65}) t^2 + \frac{a_4}{t} e^{-\frac{1}{2}(\frac{\lg t - 1}{\sigma})^2}$ (σ=0.6) (**LM1**) ⊕

3. Logarithm-Model II (Guo & Swalve, 1996):
   $y_t = a_1 + a_2 t + a_3 t^2 + \frac{a_4}{t} e^{-\frac{1}{2}(\frac{\lg t - 1}{\sigma})^2}$ (σ=0.55) (**LM2**) ⊕

4. Mixed-Log-Model IV : $y_t = a_1 + a_2 \sqrt{t} + a_3 \ln t + a_4 \sin(\frac{t}{c}) t^2 (c = 65)$ (**MIL4**) ⊕

5. Mixed-Log-Model V: $y_t = a_1 + a_2 \sqrt{t} + a_3 \ln t + a_4 \sin(\frac{t}{c}) t (c = 70)$ (**MIL5**) ⊕

III. Models with 5 parameters

1. Ali & Schaeffer(1987):
   $y_t = a_1 + a_2 (\frac{t}{c}) + a_3 (\frac{t}{c})^2 + a_4 \ln(\frac{c}{t}) + a_5 \ln^2(\frac{c}{t})$ (c=305) (**AS**) ⊕

2. Modified W-II (Guo & Swalve, 1995):
   $y_t = a_1 + a_2 t + a_3 \sin(\frac{t}{c}) t^2 + a_4 \sin(\frac{t}{c}) t^3 + a_5 e^{-0.055t} (c = 100)$ (**MW2**) ⊕

3. Modified KH-I (Guo & Swalve, 1995):
   $y_t = a_1 + a_2 t + a_3 t^2 + a_4 t^3 + a_5 \ln t$ (**MK1**) ⊕

4. NEW5 : $y_t = a_1 + a_2 t + a_3 t^2 + a_4 t^3 + a_5 \sqrt{t}$ (**NEW5**) ⊕

5. NEW6 : $y_t = a_1 + a_2 \sqrt{t} + a_3 t + a_4 t^3 + a_5 t^{\lg t}$ (**NEW6**)

6. Mixed-Log-Model VI:
   $y_t = a_1 + a_2 \sqrt{t} + a_3 \ln t + a_4 \sin(\frac{t}{c}) t + a_5 \sin(\frac{t}{c}) t^2 (c = 100)$ (**MIL6**) ⊕

7. Mixed-Log-Model VII:
   $y_t = a_1 + a_2 \ln t + a_3 \tanh(\lg t) t + a_4 \sin(\frac{t}{c}) t^2 + a_5 \sin(\frac{t}{c}) t^3 (c = 80)$ (**MIL7**)

⊕ denotes a model also used for the estimation of variance components in the second part of the study