A Comparison of Fixed and Random Regression Models Applied to Dairy Test Day Production Data

Z. Liu, J. Jamrozik and G. Jansen

Canadian Dairy Network, 150 Research Lane, Suite 307, Guelph, Ontario, Canada N1G 4T2

Introduction

The fixed regression model (FRM) developed in the early stages of modelling test day (TD) data (Ptak and Schaeffer, 1993) was a simplified version of the random regression model (RRM) (Jamrozik et al., 1997). It is assumed in the FRM that there are constant animal and permanent environmental effects throughout the course of the lactation. The heterogeneous variances at different lactation stages are not taken into account and correlations between any two tests are assumed to be equal. Although cows belonging to different sub-groups, eg. age-parity-season classes, are allowed to have different average lactation curves, deviations from the average curves are not modelled by the FRM. In the RRM, fixed lactation curves model means of data over stages of lactation while random lactation curves model the variance-covariance structure of TD data.

Little work has been done to compare the fixed and random regression models used in animal breeding. The objective of this study was to compare the two models using Canadian Holstein TD data.

Materials and Methods

In Canada, dairy TD data have been recorded and archived only since 1988. Thus, any lactations initiated before this point were discarded. Only the first three lactations of Holstein cows were selected for this study. It was required that the first test of any lactation be before 90 days in milk (DIM) and only tests between 5 and 305 DIM were used. The set of programs from the national somatic cell score evaluations was used for the FRM analysis (Reents et al., 1995) whereas a new set of programs were developed for the RRM evaluation (Jamrozik et al., 1997). Table 1 summarizes the differences in data used in the RRM and FRM evaluations, after editing on pedigree information, calving dates, lactation numbers, and age at calving. Although these differences existed, it is reasonable to assume that the two datasets were essentially the same.

Table 1. Statistics of TD data for the Holstein breed

model	TD records	cows with records	animals in total	herd-test-day-parity classes
RRM	11,544,946	786,894	1,520,096	1,219,679
FRM	11,395,391	820,092	1,235,823	1,213,380

Aside from the random regressions for animal genetic and permanent environmental effects, there were several other differences between the models. Fixed lactation curves were fitted within 152 Region-Age-Parity-Season classes (RAPS) in the RRM, and within 24 Age-Parity-Season classes (APS) in the FRM. There were minor differences in the definitions of genetic groups for unknown parents. The FRM analysed only protein yield in lactations 1 to 3 whereas the RRM evaluation also included milk yield, fat yield and somatic cell scores in a multiple trait analysis. For both models, EBV's for protein yield in three lactations on a 305-day basis were combined into a single EBV using equal weighting factors (1/3).

EBV's from the 305-day animal model (AM305)

currently used in Canadian national evaluations were also compared with FRM and RRM EBV's. Heterogeneous herd variances were accounted for in the AM305 evaluation, whereas RRM and FRM evaluations used raw TD data without adjusting for heterogeneous herd variances.

Residual analysis was conducted to check the goodness of fit of the models. Several models were fitted to the residuals, including APS or RAPS classes for the FRM and RRM, respectively, and quadratic regressions on DIM. Residuals were also plotted against DIM.

Results and Discussions

1. Analyses of Estimated Breeding Values

Table 2 shows means, standard deviations and correlations of EBV for common cows between any two of the three evaluations. Tables 3 and 4 give the same statistics for bulls with different number of daughters. As the TD and 305-day datasets cover different time periods, RRM or FRM EBV's had different averages than those of the AM305 evaluation. Different definitions of phantom parent groups may partially explain the small differences in averages of FRM and RRM EBV's. As a result of model differences, the RRM had larger variances of EBV than the FRM, both for cows and bulls. It is interesting to note that RRM EBV's were more highly correlated with EBV's of AM305 than with FRM EBV's for cows with records or bulls with daughters' records, although for animals without records the correlations between RRM and FRM were higher than between RRM and AM305. This result may indicate a problem associated with FRM in modelling TD data.

It was found across all pairs of models that correlations of bull EBV's increased with number of daughters in evaluations as expected, except that there was a decrease in correlations between the FRM and RRM for bulls with 20 to 100 daughters. For bulls with more than 1000 daughters, the highest correlation (0.98) was found between the RRM and AM305, while the lowest (0.92) was between the RRM and FRM.

2. Residual Analyses

The first part of the residual analysis concerned the goodness of fit of the fixed lactation curves. The small R^2 values (Table 5) of 0.0001 and 0.03 for model [1] in the RRM and FRM, respectively, indicate that the fixed regression curves on DIM successfully modelled the overall effects of APS or RAPS in the TD data.

In the second part of the residual analysis, an attempt was made to examine residuals within APS or RAPS classes. When model [2] is compared to model [1], the difference in the R^2 values indicates whether deviations of cows' individual lactation curves from their corresponding fixed average lactation curve were significant in the residuals. For the RRM, the two R^2 values did not differ; however, fitting an additional quadratic function on DIM within APS classes significantly increased the R^2 value for FRM residuals. This result shows that the FRM did not model the variation within APS properly.

The third part of residual analysis tried to quantify the overall effects of DIM on the residuals of both FRM and RRM. In the last two rows of Table 4, neither models [3] nor model [4] explained much variation in the RRM residuals, but the FRM had an R^2 of 0.11 for both models. This suggests once again that FRM did not remove DIM effects properly. Figure 1 shows average residuals at each DIM, by lactation. The RRM residuals were close to zero at all DIM and showed very little pattern in relation to DIM whereas a clear pattern of negative residuals was shown in later stages of lactation for the FRM.

Based on the results shown above, RRM gives a reasonable goodness of fit to TD data and demonstrates superiority over the FRM in modelling TD data. It is, therefore, recommended that the RRM rather than the FRM should be used in routine genetic evaluation of TD production traits in dairy cattle. The FRM may still be an efficient method to study standard lactation curves for groups of animals.

Literature cited

Jamrozik, J., Schaeffer, L.R., Liu, Z. & Jansen, G. 1997. Multiple trait random regression TD model for production traits. *Interbull Bulletin No.16, 43.*

Ptak, E. & Schaeffer, L.R. 1993. Use of TD yields for genetic evaluation of dairy sires and cows. *Livest. Prod. Sci. 34*, 23.

Reents, R., Dekkers, J.C.M. & Schaeffer, L.R. 1995. Genetic evaluation for somatic cell score with a TD model for multiple lactations. *J. Dairy Sci. 77*, 2671.

Table 2. Comparison of RRM, FRM and AM305 cow EBV's for protein yield

	All cows				Cows with at least 2 TD records and/or an official 305-day proof			
	Cows	Mean	S.D.	Corr.	Cows	Mean	S.D.	Corr.
RRM	1091045	-13.1	20.4	0.89	720668	-7.7	20.7	0.89
FRM		-10.9	18.5			-6.5	18.8	
RRM	1488147	-17.2	19.9	0.86	778679	-8.0	20.6	0.90
AM305		-15.5	22.9			-4.3	19.9	
FRM	1110627	-11.0	18.4	0.82	722840	-6.5	18.8	0.86
AM305		-9.8	21.2			-4.1	19.9	

Table 3. Comparison of RRM, FRM and AM305 bull EBV's for protein yield

	All bulls				Bulls with >1000 daughters			
	Bulls	Mean	S.D.	Corr.	Bulls	Mean	S.D.	Corr.
RRM	23,861	-19.7	20.9	0.85	115	3.0	25.2	0.92
FRM		-20.4	20.2			2.5	20.0	
RRM	31,495	-22.3	19.1	0.84	115	3.0	25.2	0.98
AM305		-24.6	24.7			7.3	23.3	
FRM	25,233	-20.6	20.1	0.79	114	4.0	20.3	0.93
AM305		-20.3	25.2			9.0	23.3	

Table 4. Comparison of RRM and FRM bull EBV's with different numbers of daughters

Daughters	Bulls	RRM		Corr.	FRM	
		Mean	S.D.		Mean	S.D.
0	10,577	-29.6	13.0	0.78	-29.4	13.9
1 - 20	9,549	-16.8	19.2	0.81	-17.4	19.4
21 - 50	1,866	-2.6	24.4	0.74	-5.5	23.1
51 - 100	1,320	4.9	24.8	0.81	0.5	22.0
101 - 1000	434	2.9	27.5	0.87	-0.6	24.9
> 1000	115	3.0	25.2	0.92	2.5	20.0

Table 5. Percentage R^2 values of the models fitted to the residuals of the FRM and RRM [1]

Model of residual analysis	RRM	FRM
[1] $e_{ij} = A_i + \varepsilon_{ij}$	0.01	3
[2] $e_{ij} = A_i + b_{i1}*DIM(A_i) + b_{i2}*DIM^2(A_i) + \varepsilon_{ij}$	1	15
[3] $e_{ij} = DIM_i + \varepsilon_{ij}$	2	11
[4] $e_{ij} = \mu + b_1*DIM_i + b_2*DIM^2_i + \varepsilon_{ij}$	1	11

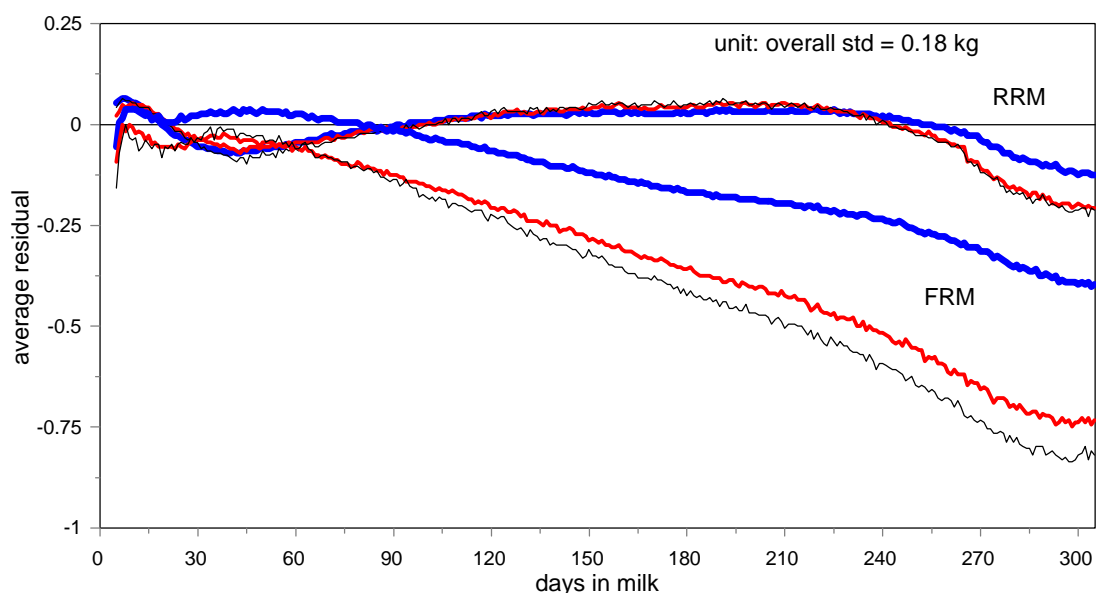[1] A_i is APS class for the FRM and RAPS class for the RRM.

Figure 1. Average RRM and FRM residuals of test day protein yields for the first three lactations (the thickest/thinnest for first/third lactation).