Res. Bull. Natl. Agric. Res. Cent. for Hokkaido Reg.193,95-101 (2011)

Running head: LEGENDRE POLYNOMIALS FOR MULTIPLE PARITIES

Determining the Order of Legendre Polynomials for Multi-parity Random Regression Test-day Animal Model

Kenji Togashi[1,*], C.Y. Lin [2], Koichi Hagiya[3], and Takatoshi Nakanishi[3]

Summary

This study aims to determine the appropriate order of Legendre polynomial (LP) for a three-parity random regression (RR) model using the Japanese Holstein production data. The goodness of fit of the RR models with second-, third-, fourth-, and fifth-order LP for the fixed and random effects was evaluated based on five statistical criteria: the three residual variances of the first three parities, the log likelihood function (-2logL), and Akaike's information criterion (AIC). A model that produces the smallest AIC indicates the best fit to the data. The percentage of reduction in residual variances with increasing order of LP was greater for the 2nd and 3rd parity than for the 1st parity, suggesting that the 1st lactation milk is more robust to the order of LP than the 2nd or the 3rd lactation milk because the first lactation curve is more persistent than the 2nd and 3rd lactation curves. All five statistical criteria decrease with increasing order of LP for the additive genetic and permanent environmental effects, suggesting that the higher the order of LP fitted for RR coefficients, the more accurate the RR model is. However, the rate of decrease for the five statistical criteria (as measured by the percentage of reduction) decreases as the order of LP goes beyond 4. When the order of LP for the fixed regression coefficients increases from 2 to 4, there are little changes in the five statistical criteria, indicating that the fitting of a RR model with a second-order LP for the fixed regression coefficients is justified. Based on the results of this study, it is advisable to fit a RR model with a fourth-order LP for the random effects and a second-order LP for the fixed effect in genetic evaluation of the milk EBV of the first three parities.

Key words: random regression, Legendre polynomials, multiple lactations

1) National Agricultural Research Centre in Hokkaido Region, Hitsujigaoka 1, Toyohiraku, Sapporo, Japan 0628555
2) Dairy and Swine Research and Development Centre, Agriculture and Agri-Food Canada, 2000 College St., P.O. Box 90-STN Lennoxville, Sherbrooke, QC, Canada J1M 1Z3
3) National Livestock Breeding Institute, Nishigo-mura, Nishishirakawa-gun, Fukushima, Japan 9618511
Correspondence
K. Togashi
National Agricultural Research Centre in Hokkaido Region
Hitsujigaoka 1, Toyohira-ku, Sapporo, Japan 0628555
Tel.: +81-11-8579263; Fax: +81-11-8592178.
E-mail: tkenji@naro.affrc.go.jp

Introduction

The use of random regression (RR) test-day animal model for genetic evaluation is appealing because different lactation curves of cows and environmental effects associated with each individual test days can be better accounted for (Schaeffer and Dekkers, 1994; Jamrozik et al., 1997). Various functions have been used to describe the distribution of the test day yields across lactation (i.e., lactation curve) for a RR model (e.g., Wood, 1967; Wilmink, 1987; Ali and Schaeffer, 1987; Kirkpatrick et al., 1990). Of these functions, Legendre polynomial (LP) for a RR model (Kirkpartick et al., 1990) was widely used for genetic evaluation of dairy cattle. At present, many countries have used the RR test day model for national genetic evaluation (Interbull, 2009). Jamrozik et al (1996) reported that a linear function of 5 covariates and a linear function of 3 covariates produced small practical difference in terms of variance and covariance components, EBV and prediction errors. Misztal et al. (2000) reported that the estimates of genetic parameters can vary with varying orders of LP. Jamrozik and Schaeffer (2002) found that the test day model using Legendre polynomials outperformed that using the lactation curve function with the same number of parameters. Norberg et al. (2004) reported that a RR model with a fourth-order LP is the best model for determining genetic parameters for test-day electrical conductivity of first-lactation milk. Liu et al. (2006) compared different orders of LP in RR models for first-lactation milk yield and found that the choice of the order of the LP for a RR model varied depending upon the statistical criterion used for model comparison.

The studies of the optimal order of LP in a RR models are well documented for the first lactation, but are quite lacking for multiple lactations. Currently, the Japanese national genetic evaluation is based on a lactation model and plans to switch to a three-lactation RR animal model in the near future. It is, therefore, necessary to determine an appropriate number of covariates to describe the lactation shapes of the first three lactations. This study aimed to compare the performance of different RR models with quadratic, cubic, quartic, and quintic Legendre polynomials using the first three-parity milk production data of the Japanese Holsteins.

Materials and methods

Detail description of data and statistical models used for this study were given in Togashi et al. (2008). Briefly, data consisted of Japanese Holstein TD milk yields of the first three lactations from 1996 to 2000. All lactations of the first three parities were required to have a minimum of 10 TD records and the number of cows per herd-test-date is required to have a minimum of 7. The TD records of milk yields within and between parities are treated as separate traits. The additive genetic effects and permanent environmental effects of the test day yields were modeled using a quadratic, cubic, quartic or quintic Legendre polynomial (LP) which corresponds to the fitting of 3, 4, 5, or 6 covariates, respectively. The three-parity test day animal model used was as follows:

$$y_{ijkl/s} = \mathrm{HTD}_{i/s} + \sum_{q=0}^{x} f_{jq/s} z_{qd/s} + \sum_{q=0}^{x} p_{kq/s} z_{qd/s} + \sum_{q=0}^{x} a_{kq/s} z_{qd/s} + e_{ijkl/s}$$

where $y_{ijklq/s}$ = test-day milk record in the sth parity,

$\mathrm{HTD}_{i/s}$ = the i^{th} herd-TD effect in the sth parity (s = 1, 2, or 3),

$z_{qd/s}$=LP coefficients (covariates) for DIM d corresponding to polynomial q in the sth parity,

$f_{jq/s}$ = the fixed regression coefficient for the sth parity specific to j^{th} location-age-season subclass (two locations, 3 ages and 4 seasons of calving gives a total of 24 subclasses per parity) evaluated at DIM d corresponding to Legendre polynomial q,

$p_{kq/s}$ = the permanent environmental RR coefficient of animal k corresponding to polynomial q in the s[th] parity,

$a_{kq/s}$ = the additive genetic RR coefficient of animal k corresponding to polynomial q in the s[th] parity,

$e_{ijkl/s}$ = the residual effect for each test-day yield.

x = the order of LP where x is the second, third, fourth or fifth order.

The residual effects were assumed uncorrelated within and between animals. The computer program REMLF90 (Misztal et al, 2000) was used to obtain the (co)variance components.

The goodness of fit of the different test-day RR models with varying orders of LP for modelling the fixed regression effect of location-age-season class and the random effects of both additive genetic and permanent environment was evaluated based on the residual variances (RV) of the first three parities, the log likelihood function (-2logL), and Akaike's information criterion (AIC) (Akaike, 1969; 1973).

Results and discussion

The Goodness of Fit of the RR Models with Varying Orders of LP for the Fixed Effect

As suggested in literature and supported by the results of this study explained in the subsequent section, the fourth-order Legendre polynomial for the random regression coefficients of both the additive genetic and permanent environmental effects was found to be an adequate fit. Consequently, the effects of changing the order of LP from 2 to 4 for the fixed effect (i.e., location-age-season) on the statistical criteria were investigated with the order of fit for the random regression coefficients being fixed at 4. Table 1 shows the residual variances (RV) of the first three parities, log likelihood function (-2logL) and Akaike's information criterion (AIC) when the order of LP for fitting the fixed regression effect within location-age-season subclass varies from 2 to 4. The results indicate that the statistical criteria compared are pretty robust to the changes in the order of LP for the fixed regression effect within location-age-season subclass. Thus, the fitting of a second- or third-order of LP for the fixed location-age-season effect would suffice. Subsequently, the LP for the fixed location-age-season effect was fitted with the third order for the purpose of studying the effects, on the statistical criteria, of fitting different orders of LP for the random regression coefficients of the random effects. Ideally, all possible combinations between the order of LP for both the fixed and random effects of the test-day model should be examined, but this would involve a lot more computations. As a compromise, this study assesses the adequate order of fit for the fixed effect first, followed by studying the adequate order of fit for both random effects (genetic and permanent environmental effects).

The Goodness of Fit of the RR Models with Varying Orders of LP for Random Effects

This study applied equal order of LP for both genetic and permanent environmental effects as suggested in literature (Olori et al. 1999; Pool and Meuwisson, 2000; Pool et al., 2000). Table 2 shows the changes in the statistical criteria (RV of the first three parities, -2logL and AIC) when both additive genetic and permanent environment effects were fitted with equal order of LP (2-2, 3-3, 4-4, or 5-5). The residual variances of the first three parities all decrease as the order of LP increases from 2 to 5. The decrease in residual variances was greater for the second and third parity than for the first parity. This means that the genetic evaluation of the first-lactation milk EBV using a test-day model is more robust to the varying order of LP for the random effects than the evaluation of the second- or third lactation milk EBV because the first lactation curve is more persistent than the 2nd and 3rd lactation curve (Muir et al., 2004; Togashi et al., 2007). Therefore, it appears justified to fit a higher order of LP for the second and third parity than for the first parity in multi-parity analysis.

The value of AIC decreases as the order of LP for the additive and permanent environmental effects increases from 2 to 5. The order of LP that produces the smallest value of AIC is considered to best fit the data among the orders of LP studied. The statistics -2logL increase slightly with the increasing order of LP. The percentage of reduction (PR) in RV, -2logL and AIC due to fitting a higher order of LP for the additive and permanent environmental effect is given in Table 3. The percentage of reduction for all statistical criteria increases as the order of LP increases up to 4 and then decreases. The results from Tables 2 and 3 combine to suggest that the fitting of a RR model with a fourth-order LP for both additive genetic and permanent environmental effects is justified. The current Canadian genetic evaluation is based on a multiple-trait RR model with a fifth-order LP (6 covariates) for the first three lactations (Canadian Dairy Network, 2009).

Based on the results of this study, the fitting of a RR Model with a fourth-order LP for both the additive genetic and permanent environmental RR coefficients and a second-order LP for the fixed regression coefficients of the fixed effect is recommended in terms of both the computational demand and evaluation accuracy for the first three-lactation milk EBV.

Conclusions

The estimation of milk EBV for the 1st parity is more robust to varying orders of Legendre polynomial (LP) than for the 2nd or the 3rd milk EBV because the 1st parity lactation curve is more persistent than the 2nd or the 3rd parity lactation curve. The changes in the statistical criteria, particularly in both the log likelihood function (-2logL) and Akaike's information criterion (AIC) stabilize when the random regression model was fitted with a fourth-order LP (5 covariates) for the additive genetic and permanent environmental effects. The five statistical criteria showed little changes with LP of orders increasing from 2 to 4 for the fixed regression effect. Considering both the computational demand and the accuracy of evaluation, the fitting of a random regression animal model with a fourth-order LP for the random effects and a second-order of LP for the fixed effect is recommended for Japanese national genetic evaluation of the first three-lactation milk production.

References

Akaike, H. (1969) Fitting autoregressive models for prediction. *Annals of the Institute of Stat. Math,* **21**,243-247.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Int. Symp. Information Theory,* Pages 267-281. Budapest, Hungary.

Ali, T. E., and L. R. Schaeffer. (1987) Accounting for covariance among test day milk yields in dairy cows. *Can. J. Anim. Sci.*, **67,**637-644.

Canadian Dairy Network. (2009) Available at.

Interbull. (2009) Available at http://www-interbull.slu.se/eval/framesida-prod.htm.

Jamrozik, J., G. J. Kistemaker, J. C. M. Dekkers, and L. R. Schaeffer. (1996) Comparison of possible covariates for use in a random regression model for analysis of test day yields. *J. Dairy Sci.*, **80,**2550-2556.

Jamrozik, J., and L. R. Schaeffer. (1997) Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins. *J. Dairy Sci.* ,**80**,762-770.

Jamrozik, J., L. R. Schaeffer, and F. Grignola. (1998) Genetic parameters for production traits and somatic cell score of Canadian Holsteins with multiple trait random regression model. *Proc. of the 6th WCGALP*, **23**,303-306.

Jamrozik, J., and L.R. Schaeffer. (2002) Bayesian comparison of random regression models for test-day yields in dairy cattle. *Proc. 7th WCGALP*. CD-ROM Commu. # 01-03. Montpellier, France.

Kirkpatrick, M., D. Lofsvold, and M. Bulmer. (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*, **124**,979-993.

Liu, Y. X., J. Zhang, L. R. Schaeffer, R. Q. Yang, and W. L. Zhang. (2006) Optimal random regression models for milk production in dairy cattle. *J. Dairy Sci.*, **89**,2233-2235.

Misztal, I., T. Strabel, J. jamrozik, J.E.A. Mantysaari, and T.H. Meuwissen. (2000) Strategies for estimating the parameters needed for different test say models. *J. Dairy Sci.*, **83,**1125-1134.

Muir, B. L., J. Fatehi, and L. R. Schaeffer. (2004) Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holstein. *J. Dairy Sci.*, **87,**3029-3037.

Norberg, E., G.. W. Rogers, R. C. Goodling, J. B. Cooper, and P. Madsen. (2004) Genetic parameters for test-day electrical conductivity of milk for first-lactation cows from random regression models. *J. Dairy Sci.* ,**87,**1917-1924.

Olori, V. E., W. G. Hill, B. J. McGuirk, and S. Brotherstone. (1999) Estimating variance components for test day milk records by restricted maximum likelihood with a random regression animal model. *Livest. Prod. Sci.* ,**61,**53-63.

Pool, M. H., and T.H.E. Meuwissen. (2000) Prediction of daily milk yields from a limited number of test days using test day models. *J. Dairy Sci.*, **82,**1555-1564.

Pool, M. H., L. L. G. Janss, and T. H. E. Meuwissen. (2000) Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.*, **83,**2640-2649.

Schaeffer, L. R, and J. C. M. Dekkers. 1994. Random regression in animal models for test-day production in dairy cattle. *Proc. 5th WCGALP* **18,** 443-446.

Togashi, K., and C. Y. Lin. (2008) Genetic improvement of total milk yield and total persistency of the first three lactations in dairy cattle. *J. Dairy Sci.*, **91,**2836-2843.

Togashi, K., C.Y. Lin, Y. Atagi, K. Hagiga, J. Sato, and T. Nakanishi. (2008) Genetic characteristics of Japanese Holstein cows based on multiple-lactation random regression test-day animal models. *Livestock Sci.*, **114,**194-201.

Wilmink, J. B. M. (1987) Adjustment of test day milk, fat and protein yield for age, season and stage of lactation. *Livest. Prod. Sci.*, **16**,335-348.

Wood, P. D. P. (1967) Algebraic model of the lactation curve in cattle. *Nature* **216,**164-165.

Table 1. Residual variances (RV) of the first three parities, log likelihood function (-2logL) and Akaike's information criterion (AIC) for random regression models fitted with 2 to 4 orders of LP for fixed effect and with LP of order 4 for both additive and permanent environmental effects

Criteria	2 order	3 order	4 order
RV of 1^{st} parity	4.090	4.086	4.097
RV of 2^{nd} parity	5.981	5.986	5.977
RV of 3^{rd} parity	8.037	8.030	8.033
-2logL	485687.37	485749.60	485762.55
AIC[1]	486293.37	486355.60	486368.55

Table 2. Residual variances (RV) of the first three parities, log likelihood function (-2logL) and Akaike's information criterion (AIC) when both additive and permanent environment effects were fitted with the same order of LP (2-2, 3-3, 4-4, or 5-5)

Statistical criteria	Order 2-2	Order 3-3	Order 4-4	Order 5-5
RV of 1st parity	4.806	4.428	4.086	3.918
RV of 2nd parity	7.547	6.744	5.986	5.598
RV of 3rd parity	10.220	9.121	8.030	7.450
-2logL	491727.1	489831.8	485749.6	484076.9
AIC[1]	491949.1	490053.8	486355.6	484946.9

Table 3. Percentage of reduction (PR) in residual variances (RV), log likelihood function (-2logL) and Akaike's information criterion (AIC) by fitting higher order of LP

	Percentage of reduction (PR)		
Statistical criteria	From order 2 to 3	From order 3 to 4	From order 4 to 5
RV of 1st parity	7.9	7.7	4.1
RV of 2nd parity	10.6	11.2	6.4
RV of 3rd parity	10.8	12.0	7.2
-2logL	0.4	0.8	0.3
AIC	0.4	0.8	0.1

PR=$100\times(C_i - C_{i+1})/C_i$ where C refers to a given statistical criterion (RV, -2logL, or AIC) and the subscript i is equal to order 2, 3, 4 or 5.