



Impact of the Order of Legendre Polynomials in Random Regression Model on Genetic Evaluation for Milk Yield in Dairy Cattle Population

Jianbin Li^{1*†}, Hongding Gao^{2†}, Per Madsen², Rongling Li¹, Wenhao Liu¹, Peng Bao³, Guanghui Xue³, Yundong Gao¹, Xueke Di⁴ and Guosheng Su^{2*}

¹ Dairy Cattle Research Center, Shandong Academy of Agricultural Sciences, Jinan, China, ² Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark, ³ Shandong OX Livestock Breeding Industry Co., Ltd, Jinan, China, ⁴ Linqing Rutai Animal Husbandry Co., Ltd, Liaocheng, China

OPEN ACCESS

Edited by:

Mariza De Andrade,
Mayo Clinic, United States

Reviewed by:

Hong Zhang,
Merck, United States
Zhaoxia Yu,
University of California, Irvine,
United States

*Correspondence:

Jianbin Li
msdljb@163.com
Guosheng Su
guosheng.su@qgg.au.dk

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Statistical Genetics and Methodology,
a section of the journal
Frontiers in Genetics

Received: 22 July 2020

Accepted: 24 September 2020

Published: 05 November 2020

Citation:

Li J, Gao H, Madsen P, Li R,
Liu W, Bao P, Xue G, Gao Y, Di X and
Su G (2020) Impact of the Order
of Legendre Polynomials in Random
Regression Model on Genetic
Evaluation for Milk Yield in Dairy Cattle
Population. *Front. Genet.* 11:586155.
doi: 10.3389/fgene.2020.586155

The random regression test-day model has become the most commonly adopted model for routine genetic evaluations in dairy populations, which allows accurately accounting for genetic and environmental effects over lactation. The objective of this study was to explore appropriate random regression test-day models for genetic evaluation of milk yield in a Holstein population with a relatively small size, which is the common situation in regional or independent breeding companies to perform genetic evaluation. Data included 419,567 test-day records from 54,417 cows from the first lactation. Variance components and breeding values were estimated using a random regression test-day model with different orders (from first to fifth) of Legendre polynomials (LP) and accounted for homogeneous or heterogeneous residual variance across the lactation. Models were compared based on Akaike information criterion (AIC), Bayesian information criterion (BIC), and predictive ability. In general, models with a higher order of LP showed better goodness of fit based on AIC and BIC values. However, models with third, fourth, and fifth order of LP led to similar estimates of genetic parameters and predictive ability. Models with assumption of heterogeneous residual variances achieved better goodness of fit but yielded similar predictive ability compared with those with assumption of homogeneous residual variances. Therefore, a random regression model with third order of LP is suggested to be an appropriate model for genetic evaluation of milk yield in local Chinese Holstein populations.

Keywords: genetic evaluation, genetic parameters, random regression, test-day records, legendre polynomials

INTRODUCTION

The random regression test-day model has been widely used in genetic evaluation for production traits in dairy cattle and has many advantages, including more accurately accounting for genetic and environment effects at different stages of lactation, thus resulting in more reliable genetic evaluation (Schaeffer and Dekkers, 1994; Jamrozik et al., 1997b; Schaeffer et al., 2000; Jensen, 2001; Schaeffer, 2004). The test-day model is significantly better than the lactation model (using full and extended 305 days lactation records) with 2–3% increase in accuracy for bulls and 6–8% for cows for milk yield first lactation (Kistemaker, 1997). In addition, the test-day model allows predicting estimated breeding value (EBV) for each test-day, each particular period, or the full lactation (305 days) and persistency (Jamrozik et al., 1997b).

The functions generally used to model the lactation curve include Woods' model (Wood, 1967), Wilmink's function (Wilmink, 1987), spline function (White et al., 1999), and Legendre polynomial (LP) function (Kirkpatrick et al., 1990). Because of differences in production environments and management systems, optimal functions for test-day models in different countries may be different (Reinhardt et al., 2002; Mrode et al., 2003). Several studies have shown that LPs performed well in random regression test-day models, but there is no "gold standard" reported in literature on choosing optimal order of LPs in the model, and the choice of the order of fit is highly dependent on the practical data structures and populations. For example, a fourth order LP (LP4) was used for national genetic evaluations in Canada and Italy, and a fifth order LP (LP5) was used in the United Kingdom (Muir et al., 2007). The joint Nordic test-day model was a multivariate model for milk, protein, and fat in lactation one to three (in total, nine traits), the genetic and permanent environmental effects were modeled by second order LP and an exponential term $e^{-0.04 \times DIM}$ (Lidauer et al., 2015).

The dairy herd improvement (DHI) project was first implemented in four provinces in the 1990s in China. In 2006, the Ministry of Agriculture of China approved a project to promote the DHI project in eight provinces where there were many large dairy populations (Miglior et al., 2009). Recently, the project has been expanded to 25 provinces in China, where provincial DHI laboratories and data centers have been established (Yearbook, 2016), and there were about 700,000 cows recording milk production in China each year. Many organizations and companies prefer to conduct genetic evaluation based on their own populations by considering the differences among populations. In general, these populations are relatively small in size. Therefore, an appropriate parametric function that can sufficiently account for the lactation curve and avoid overparameterization is critical for relatively small populations when using a random regression test-day model to estimate genetic parameters and breeding values. Relatively little research has been placed on the impact of parametric functions on genetic evaluation in a numerically small size dairy cattle population.

The aim of this study was to investigate the impact of the order of LPs on estimation of genetic parameters and prediction of breeding values for the milk yield trait and find appropriate orders of LPs in random regression test-day models for genetic evaluation in numerically small cattle populations, such as genetic evaluations at the regional level in China.

MATERIALS AND METHODS

Data

Data were obtained from the database at the Dairy Cattle Research Centre DHI Lab, Shandong Academy of Agricultural Sciences. First were lactation records from 2004–2015 that fulfill the following criteria: ages at calving between 20 and 38 months, daily milk yield between 5 and 80 kg, days in milk (DIM) between 5 and 305, and cows with at least three test-day records

were extracted. The final data set consisted of 419,567 test-day records from 54,417 cows. The number of records in each DIM class ranged from 576 to 1768. Descriptive statistics of the data are presented in **Table 1**. The traced pedigree included 104,884 individuals.

Model

We used first to fifth order LPs (LP1–LP5) to fit random regression test-day models, respectively. The model equation was as follows:

$$y_{itklm} = HYS_l + AGE_m + DIM_t + \sum_{k=0}^{nr} z_{itk} a_{ik} + \sum_{k=0}^{np} z_{itk} p e_{ik} + e_{itklm} \tag{1}$$

where y_{itklm} is the observation of cow i measured at time t within herd-year-season (HYS) subclass l and age (AGE) subclass m ; DIM_t is the fixed effect of DIM in 301 classes from 5 to 305 days; a_{ik} and $p e_{ik}$ are the k th random regression of additive genetic and permanent environmental effects for cow i , respectively; z_{itk} is the k th order of LPs for cow i measured at DIM t , nr and np denote the order of LPs (from one to five) for additive genetic and permanent environmental effects, respectively; and e_{itklm} is the random residual. LPs in this study were defined as follows (Kirkpatrick et al., 1990):

$$LP_0(x) = \frac{1}{\sqrt{2}}, LP_1(x) = \sqrt{\frac{3}{2}}x, LP_2(x) = \sqrt{\frac{5}{2}}(\frac{3}{2}x^2 - \sqrt{\frac{1}{2}}),$$

$$LP_3(x) = \sqrt{\frac{7}{2}}(\frac{5}{2}x^3 - \sqrt{\frac{3}{2}}x), LP_4(x) =$$

$$\sqrt{\frac{9}{2}}(\frac{35}{8}x^4 - \frac{30}{8}x^2 + \frac{3}{8}), \text{ and}$$

$$LP_5(x) = \sqrt{\frac{11}{2}}(\frac{63}{8}x^5 - \frac{70}{8}x^3 + \frac{15}{8}x).$$

TABLE 1 | Number of records (N), mean, and standard deviation (SD) of milk production in each DIM class.

DIM class	N	Mean (kg)	SD (kg)
5–34	28,704	23.7	7.45
35–64	37,991	26.6	7.79
65–94	41,787	26.5	7.92
95–124	43,506	26.0	7.89
125–154	44,033	25.1	7.89
155–184	44,781	24.4	7.93
185–214	45,873	23.7	7.99
215–244	45,688	22.9	8.04
245–274	44,373	22.0	8.10
275–305	42,831	21.4	8.20
Total	419,567	24.2	8.13

where x is the standardized DIM, calculated as $x = \frac{2(DIM - DIM_{min})}{(DIM_{max} - DIM_{min})} - 1$, and DIM_{min} and DIM_{max} are the minimum and maximum DIM, respectively.

In this study, March, April, May, September, and October were defined as calving season one; June, July, and August as calving season two; and November, December, January, and February as calving season three, and there were 1891 herd-year-season classes in total. Calving age was classified into four levels: 20–23, 24–27, 28–31, 32 months or later. Residual variance was assumed to be either homogeneous or heterogeneous across lactation. For models with heterogeneous residual variances, residuals were divided into 10 classes (5–34, 35–64, 65–94, 95–124, 125–154, 155–184, 185–214, 215–244, 245–274, and 275–305 DIM) (Pereira et al., 2013). The heterogeneous residual variances were handled by putting different weights on residual variance for different periods of DIM. The weights were calculated as $w_i = \frac{\bar{v}}{v_i}$, where v_i is the residual variance of the i th DIM class (Table 2), and \bar{v} is the mean of residual variances.

Additive genetic variance for a particular DIM was calculated as $\sigma_{gk}^2 = z_k' G z_k$, z_k is a column vector of LP coefficients at the k th DIM, and G is covariance matrix of additive genetic effect. Permanent environmental variance for a particular DIM was $\sigma_{pk}^2 = z_k' P z_k$, matrix P is a covariance matrix of permanent environmental effect, and z_k is the same as above; EBV of a particular animal at a particular DIM was calculated as $EBV_{mk} = z_k' a$, a is column vector of additive genetic random regression coefficients of a particular animal, and z_k is same as above; the EBV for the whole lactation was calculated as $EBV_{m305} = \sum_{t=5}^{305} EBV_{mk}$. The estimation of variance components and prediction of breeding values using different models were carried out by the DMU package (Madsen et al., 2014).

Model Comparison

Models with different orders of LPs were compared using the following methods based on full and reduced data sets:

1. Akaike information criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978). AIC

was computed as $AIC = -2\log L + 2p$, where $-2\log L$ is the restricted maximum log likelihood value, p is the total number of parameters estimated in the corresponding model. Bayesian information criterion was computed as $BIC = -2\log L + p\log(n)$, where $-2\log L$ and p are the same as defined in AIC, n is the difference between number of test-day records and the rank of the fixed effects design matrix (Strabel et al., 2005). For both AIC and BIC, a lower value indicates a better goodness of fit.

2. Pearson correlation coefficient and Spearman's rank correlation coefficient between EBV of 305-day (EBV305) estimated based on data with (full data) and without (reduced data) the phenotypes of animals calving in the last year was used to assess the prediction accuracy and the similarity between the rankings of animals based on those two groups of EBV305.

RESULTS

General Statistics of TD Milk

Table 1 shows the mean for TD milk yield in the different classes of lactation, which ranged from 21.4 to 26.6 kg with standard deviations from 7.45 to 8.20 kg. An increase in milk yield was found up to 53 DIM, followed by a gradual decrease until the end of lactation. Averaged over different classes, TD milk was 24.2 kg with a standard deviation of 8.13 kg in first lactation Holstein cows.

Goodness of Fit

Table 3 presents the number of estimated parameters and model comparison based on values of AIC and BIC. Number of estimated (co)variances for random effects of models increased from 7 to 43 when increasing the order of LP from one to five. Overall, AIC and BIC tend to be in favor of more complex models with higher orders of LP. However, the differences of AIC and BIC values were smaller among LP3, LP4, and LP5 compared with those

TABLE 2 | Estimates of residual variance (v_{ij})₁ over different classes of DIM and LP when assuming heterogeneous residual variances.

DIM class	v_{1i}	v_{2i}	v_{3i}	v_{4i}	v_{5i}
5–34	29.20	18.76	13.95	14.54	15.32
35–64	17.24	15.29	15.44	14.04	12.61
65–94	14.33	14.70	13.05	11.88	11.90
95–124	14.92	14.12	12.18	12.23	11.40
125–154	15.34	12.99	12.45	11.80	11.33
155–184	15.30	12.54	12.48	11.39	11.26
185–214	14.06	12.63	11.57	11.43	10.54
215–244	12.53	12.75	11.23	11.15	10.80
245–274	11.60	11.12	10.97	9.95	9.64
275–305	16.21	10.38	8.64	9.57	9.80
Total	16.07	13.53	12.20	11.80	11.46

¹ v_{ij} is the residual variance of i th class in lactation with LP of order i .

TABLE 3 | Number of variance components estimated for additive genetic effect (nA), permanent environment effect (nPE), AIC, and BIC for model comparison.

	Order ³	nA	nPE	AIC	BIC
HO ¹	LP1	3	3	1,757,636	1,757,712
	LP2	6	6	1,739,185	1,739,327
	LP3	10	10	1,732,188	1,732,418
	LP4	15	15	1,728,916	1,729,255
	LP5	21	21	1,726,882	1,727,352
HE ²	LP1	3	3	1,725,349	1,725,426
	LP2	6	6	1,722,036	1,722,179
	LP3	10	10	1,720,665	1,720,894
	LP4	15	15	1,716,308	1,716,647
	LP5	21	21	1,710,956	1,711,426

¹HO denote assumption of homogeneous residual variance. ²HE denote assumption of heterogeneous residual variance. ³Random regression test-day model with Legendre polynomials (LP) of order i (i from 1 to 5).

among LP1, LP2, and LP3 for homogeneous residual variance. Smaller differences were observed across different orders of LP for heterogeneous residual variance compared with those for homogeneous residual variance. Furthermore, models

with assumption of heterogeneous residual variances received lower values of AIC and BIC compared with models with assumption of homogeneous residual variance, indicating better goodness of fit.

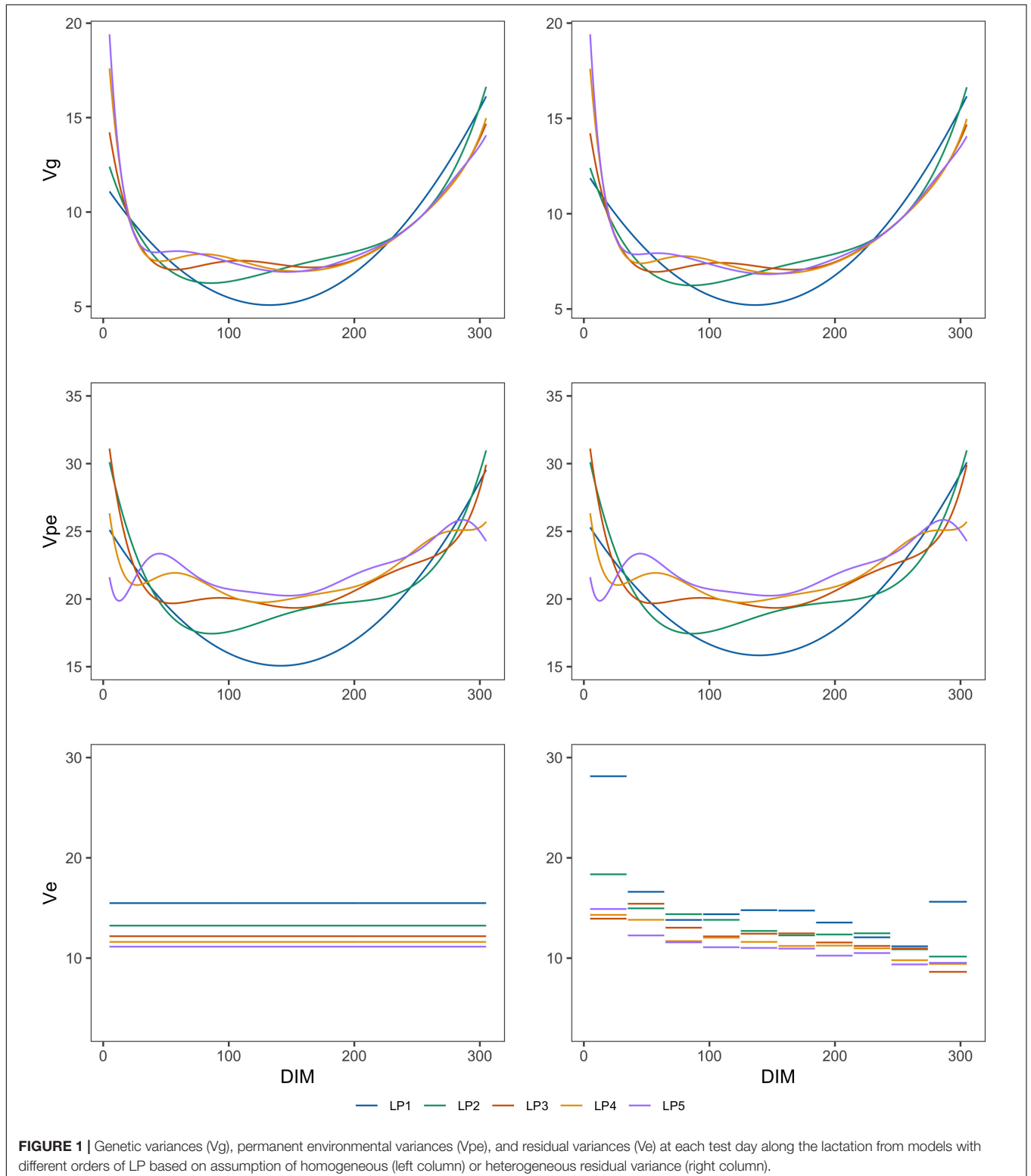


FIGURE 1 | Genetic variances (Vg), permanent environmental variances (Vpe), and residual variances (Ve) at each test day along the lactation from models with different orders of LP based on assumption of homogeneous (left column) or heterogeneous residual variance (right column).

Comparison on Estimated Variance Components and Heritability

Figures 1, 2 show the genetic variances (σ_g^2), permanent environmental variances (σ_p^2), residual variances (σ_e^2), heritabilities (h^2), and repeatabilities (Rep) calculated for each day along the lactation trajectory based on the estimated covariance function coefficients. The curve of genetic variances showed a sharp decrease in the early lactation and an increase from the middle to the end of lactation. The curve of permanent environmental variances presented similar trends as the genetic variances.

Estimates of genetic and permanent environmental variance were slightly different in models with different orders because higher order had more inflexion points. For estimates of residual variance, fitting higher order of polynomial yielded a lower horizontal line when assuming homogeneity in the model. In addition, when assuming heterogeneity in the model, although the whole horizontal line was split into 10 segments, the same trend as the homogeneous model was observed with respect to the order of fit. In particular, peaks can be seen at the beginning of lactation when fitting lower orders of polynomials. Similarly, the heritability was high at the beginning and end of lactation and lower in the middle.

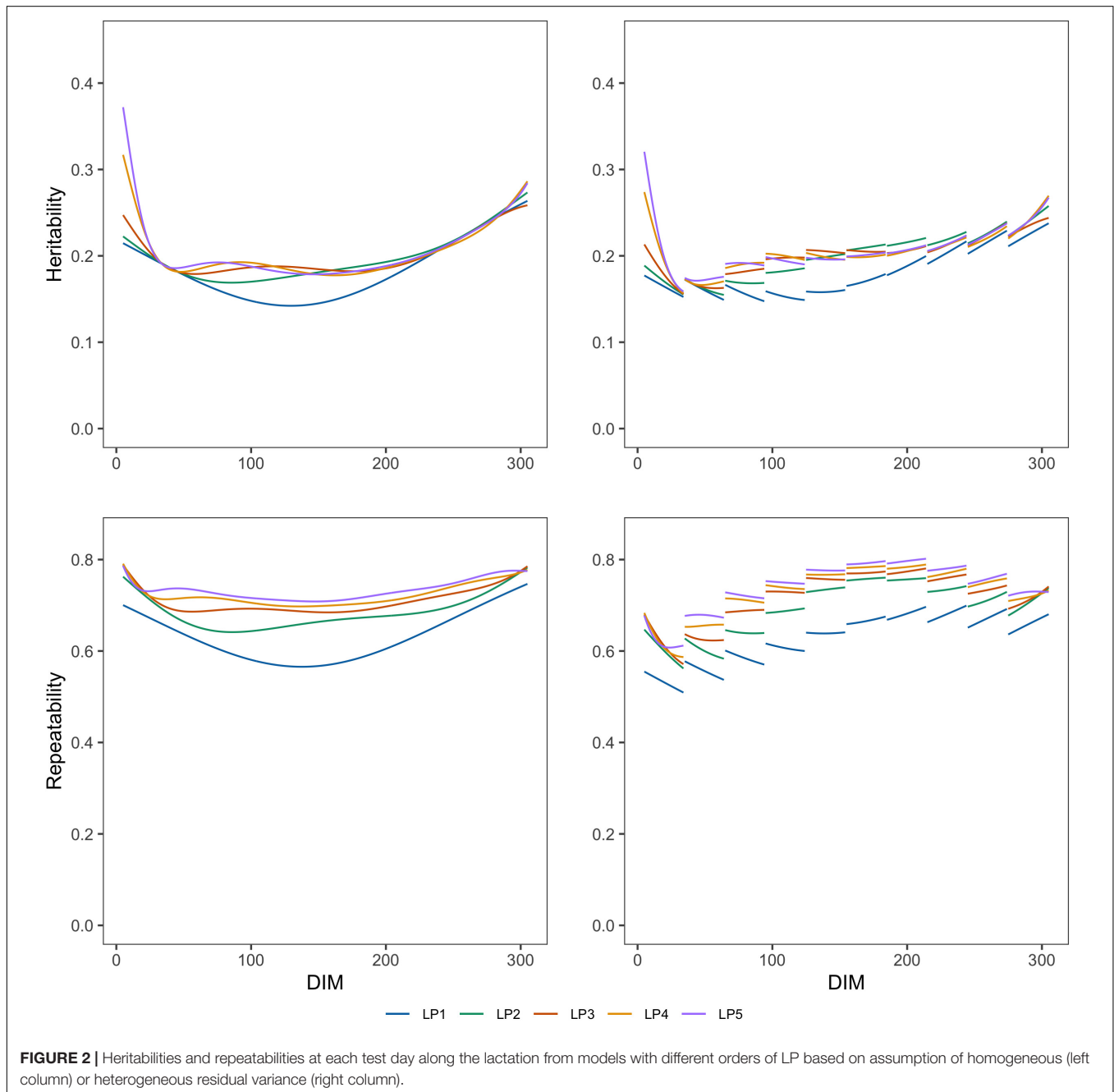


FIGURE 2 | Heritabilities and repeatabilities at each test day along the lactation from models with different orders of LP based on assumption of homogeneous (left column) or heterogeneous residual variance (right column).

TABLE 4 | Heritabilities (h^2) and Repeatabilities (Rep) for 305 days milk yield and minimal (Min) and maximal (Max) h^2 and Rep for TD in different LP considering homogeneous (HO) and heterogeneous (HE) residual variances.

	LP	HO			HE		
		305 days	Min for TD	Max for TD	305 days	Min for TD	Max for TD
h^2	LP1	0.257	0.142	0.264	0.260	0.143	0.269
	LP2	0.254	0.169	0.273	0.254	0.163	0.303
	LP3	0.250	0.179	0.259	0.250	0.163	0.304
	LP4	0.250	0.177	0.317	0.249	0.170	0.326
	LP5	0.250	0.179	0.372	0.249	0.171	0.303
Rep	LP1	0.741	0.566	0.747	0.738	0.520	0.773
	LP2	0.744	0.641	0.782	0.744	0.628	0.821
	LP3	0.748	0.685	0.788	0.748	0.639	0.821
	LP4	0.749	0.697	0.791	0.749	0.682	0.799
	LP5	0.748	0.708	0.786	0.749	0.688	0.799

TABLE 5 | Pearson correlations (Spearman's rank correlations) between EBV of 305-day milk yield obtained from full and reduced data for models with different orders of LP and with homogeneous (HO) or heterogeneous (HE) residual variance.

Residual variance	LP1	LP2	LP3	LP4	LP5
HO	0.703 (0.684)	0.713 (0.689)	0.731 (0.701)	0.730 (0.700)	0.729 (0.697)
HE	0.694 (0.682)	0.711 (0.688)	0.733 (0.702)	0.732 (0.702)	0.729 (0.698)

Repeatability decreased in early lactation and increased to the end of lactation when using the models with homogeneous residual variance although it decreased at the first stage of lactation and increased to about a DIM of 200, then decreased slightly to the end of lactation when using the models with heterogeneous variances.

Table 4 shows the heritability (h^2) and repeatability (Rep) for 305-day milk yield and the minimum and maximum values for TD milk yield. h^2 for 305 days estimated from models with different LPs ranged from 0.250 to 0.257, and Rep were from 0.741 to 0.749 when considering homogeneous variance. When considering heterogeneous variance, h^2 for 305 days were between 0.250 and 0.260, and Rep were between 0.738 and 0.749. h^2 for TD from models with different LPs ranged from 0.142 to 0.372 and Rep were between 0.566 and 0.786 when considering homogeneous variance. When considering heterogeneous residual variances, h^2 for TD ranged from 0.143 to 0.326, and Rep were between 0.520 and 0.821. Models with LP1 and LP2 led to higher estimated heritability and lower repeatability than the models with higher order. Heritability and repeatability for 305 days milk yield estimated from the models with LP3, LP4, and LP5 were very similar, ranging from 0.249 to 0.250 for heritability and from 0.748 to 0.749 for repeatability. Models with homogeneous or heterogeneous residual variances led to similar estimates of heritability and repeatability except for the LP1 model.

Comparison on Predictive Ability

Table 5 presents Pearson correlation and Spearman's rank correlation coefficients between EBV305 from full and reduced data. In general, the highest correlation was observed from LP3. For both Pearson correlations and Spearman's rank correlations, the correlation coefficients increased from LP1 to LP3 and kept

relatively stable from LP3 to LP5 for both homogeneous and heterogeneous residual variances. Models with homogeneous and heterogeneous residual variances yielded similar prediction accuracies across different orders of LP. A similar pattern was found for the rank correlations.

DISCUSSION

In this study, various criteria were used to compare random regression test-day models with different order of LPs. Different criteria for model comparison have been discussed by Druet et al. (2003) and Odegard et al. (2003). In our study, models with higher order obtained lower AIC and BIC values, which was in line with previous studies (Pereira et al., 2013). This means models with LP5 fit data best regardless of complexity. However, models with higher orders introduced more parameters and resulted in a heavier computational demand. Therefore, in practice, model selection needs to balance between goodness of fit and model complexity.

The further improvements of goodness of fit by increasing order of LP became smaller when using higher order of LP. In particular, when assuming homogeneous residual variance, there was a smaller improvement by increasing order of fit from LP3 to LP4 and from LP4 to LP5 based on values of AIC and BIC. Similar results were observed by Lopez-Romero and Carabano (2003) based on a Spanish Holstein population, in which the reduction in residual variance was small when increase order of fit from LP4 to LP6. In addition, they reported the order of LP3 was the optimum choice based on the BIC values although higher order (LP6) was needed for permanent environmental effects.

The trajectory of additive genetic variances and permanent environmental variances showed a decrease at the beginning of

lactation and an increase at the end of lactation. A similar pattern was reported by Van Der Werf et al. (1998), Pool et al. (2000), and Strabel et al. (2005) in Polish, Dutch, and Australian Holstein populations. Higher additive genetic variances at the beginning and end of lactation might be attributed to variations in the number of TD records, milk yield level, or non-genetic factors, for example, pregnancy effects (Meyer, 2005). Other studies have also shown that fitting a higher order of LP produced higher estimates of genetic variances at the edges of lactation and an oscillatory pattern along the lactation trajectory, which might be unlikely biologically (Jamrozik et al., 1997a; Pool et al., 2000; Lopez-Romero and Carabano, 2003; Meyer, 2005). This indicates that a model with a higher order (e.g., LP5) may not be more optimal than a model with a lower order (e.g., LP3 or LP4).

Heterogeneous residual variances were generally observed over the lactation when analyzing test-day records in dairy cattle populations (White et al., 1999; Brotherstone et al., 2000; Jensen, 2001). Therefore, it is reasonable to overcome this impact. It can be clearly seen that the residual variances changed across 10 DIM groups defined in this study although the differences were small when fitting higher orders of LP. More specifically, the estimated residual variances were larger in the early lactation as also shown in other studies (Swalve, 1995; Jensen, 2001). In addition, **Figure 1** reveals that the slope of the permanent environmental variance curve was affected by the change of residual variance over the lactation when assuming homogeneous residual variance in the model (**Figure 1**). This pattern can be explained by the fact that the permanent environmental effects have the ability to absorb part of the heterogeneity of the residuals (Ødegard et al., 2003).

In practical breeding programs, a critical procedure is to obtain accurate breeding values for the candidates via genetic evaluation and then make selection decisions based on the rankings of EBVs of the animals. Therefore, predictive ability is an essential property in these circumstances. In this study, predictive ability of the models was assessed by both Pearson correlations and Spearman's rank correlations between EBV305 from a full and a reduced data set to evaluate the predictive abilities between models. A model with better goodness of fit does not necessarily indicate better predictive ability, whereas a poorly fit model, such as underfitting, induces bias. This was demonstrated by the highest coefficients of Pearson and Spearman's rank correlation from models using LP of order three regardless of the assumption of residual variance. Hence, this suggests that fitting a model with order three of LPs is reasonable to achieve better predictive ability for milk yield.

CONCLUSION

The results in this study show that goodness of fit increased with increased order of LP. The results demonstrate that further improvement became small when using models with order of LP higher than three. This indicates that a minimum of LPs

of order three is necessary for estimating variance components and breeding value of milk yield. Considering a balance between goodness of fit, computational demand, and predictive ability, random regression test-day models using LP3 or LP4 appear to be the appropriate models for implementation of genetic evaluation in different Chinese Holstein populations.

AUTHOR'S NOTE

This manuscript has been released as a pre-print at bioRxiv (Li et al., 2019).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because The data are milk production of dairy cattle collected from dairy farms. Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

JL, HG, and GS conceived and designed the study. JL and HG conducted all analysis and wrote the draft. GS revised manuscript. PM maintained the DMU software package used in the statistical analysis. RL, WL, PB, GX, YG, and XD helped with the data collection and project coordination. All authors provided critical feedback and helped shape the manuscript.

FUNDING

This research was funded by Natural Science Foundation of Shandong Province (Zr2016Cm37), Key Research and Development Plan of Shandong Province (2018Gnc113003), China Agriculture Research System (Cars-36), Agricultural Scientific and Technological Innovation Project of Shandong Academy of Agricultural Sciences (Cxgc2016A04), Agricultural Improved Varieties Project of Shandong Province (2016Lzgc027), and Open Project of Technical Innovation Strategic Alliance of Livestock Breeding Industry (2019005).

ACKNOWLEDGMENTS

Discussions with Lingzhao Fang and Bingjie Li are greatly appreciated.

REFERENCES

- Akaike, H. (1973). "Information theory as an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, eds B. N. Petrov and F. Csaki (Hungary: Akademiai Kiado), 267–281.
- Brotherstone, S., White, I. M. S., and Meyer, K. (2000). Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Anim. Sci.* 70, 407–415. doi: 10.1017/s1357729800051754
- Druet, T., Jaffrezic, F., Boichard, D., and Ducrocq, V. (2003). Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.* 86, 2480–2490. doi: 10.3168/jds.s0022-0302(03)73842-9
- Jamrozik, J., Kistemaker, G. J., Dekkers, J. C. M., and Schaeffer, L. R. (1997a). Comparison of possible covariates for use in a random regression model for analyses of test day yields. *J. Dairy Sci.* 80, 2550–2556. doi: 10.3168/jds.s0022-0302(97)76210-6
- Jamrozik, J., Schaeffer, L. R., and Dekkers, J. C. M. (1997b). Genetic evaluation of dairy cattle using test day yields and random regression model. *J. Dairy Sci.* 80, 1217–1226. doi: 10.3168/jds.s0022-0302(97)76050-8
- Jensen, J. (2001). Genetic evaluation of dairy cattle using test-day models. *J. Dairy Sci.* 84, 2803–2812. doi: 10.3168/jds.s0022-0302(01)74736-4
- Kirkpatrick, M., Lofsvold, D., and Bulmer, M. (1990). Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124, 979–993.
- Kistemaker, G. J. (1997). *The comparison of random regression testday models and a 305-d model for evaluation of milk yield in dairy cattle*. Ph.D. thesis, University of Guelph, Ontario.
- Li, J., Gao, H., Madsen, P., Liu, W., Bao, P., Xue, G., et al. (2019). Comparison of different orders of Legendre polynomials in random regression model for estimation of genetic parameters and breeding values of milk yield in the Chinese Holstein population. *bioRxiv* 2019:562991.
- Lidauer, M. H., Poso, J., Pedersen, J., Lassen, J., Madsen, P., Mantysaari, E. A., et al. (2015). Across-country test-day model evaluations for Holstein. Nordic Red Cattle, and Jersey. *J. Dairy Sci.* 98, 1296–1309. doi: 10.3168/jds.2014-8307
- Lopez-Romero, P., and Carabano, W. (2003). Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein-Friesian cattle. *Liv. Prod. Sci.* 82, 81–96. doi: 10.1016/s0301-6226(03)00003-4
- Madsen, P., Su, G., Labouriau, R., and Christensen, O. (2014). "DMU—A package for analyzing multivariate mixed models," in *Paper presented 10th World Congress on Genetics Applied to Livestock Production (WCGALP)* (Vancouver: WCGALP).
- Meyer, K. (2005). Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.* 37, 473–500. doi: 10.1186/1297-9686-37-6-473
- Miglior, F., Gong, W., Wang, Y., Kistemaker, G. J., Sewalem, A., and Jamrozik, J. (2009). Genetic parameters of production traits in Chinese Holsteins using a random regression test-day model. *J. Dairy Sci.* 92, 4697–4706. doi: 10.3168/jds.2009-2212
- Mrode, R. A., Swanson, G. J. T., and Paget, M. F. (2003). "Implementation of a test day model for production traits in the UK," in *Proceedings of Interbull Meeting* (Cambridge: Cambridge University Press), 193–196.
- Muir, B. L., Kistemaker, G., Jamrozik, J., and Canavesi, F. (2007). Genetic parameters for a multiple-trait multiple-lactation random regression test-day model in Italian Holsteins. *J. Dairy Sci.* 90, 1564–1574. doi: 10.3168/jds.s0022-0302(07)71642-9
- Odegard, J., Jensen, J., Klemetsdal, G., Madsen, P., and Heringstad, B. (2003). Genetic analysis of somatic cell score in Norwegian cattle using random regression test-day models. *J. Dairy Sci.* 86, 4103–4114. doi: 10.3168/jds.s0022-0302(03)74024-7
- Odegard, J., Jensen, J., Klemetsdal, G., Madsen, P., and Heringstad, B. (2003). Genetic analysis of somatic cell score in Norwegian cattle using random regression test-day models. *J. Dairy Sci.* 86, 4103–4114.
- Pereira, R. J., Bignardi, A. B., El Faro, L., Verneque, R. S., Vercesi, A. E., and Albuquerque, L. G. (2013). Random regression models using Legendre polynomials or linear splines for test-day milk yield of dairy Gyr (*Bos indicus*) cattle. *J. Dairy Sci.* 96, 565–574. doi: 10.3168/jds.2011-5051
- Pool, M. H., Janss, L. L. G., and Meuwissen, T. H. E. (2000). Genetic parameters of Legendre polynomials for first parity lactation curves. *J. Dairy Sci.* 83, 2640–2649. doi: 10.3168/jds.s0022-0302(00)75157-5
- Reinhardt, F., Liu, Z., Bünger, A., Dopp, L., and Reents, R. (2002). "Impact of application of a random regression test day model to production trait genetic evaluations in dairy cattle," in *Proceedings of Interbull Meeting* (Cambridge: Cambridge University Press), 103–108.
- Schaeffer, L. R. (2004). Application of random regression models in animal breeding. *Livest. Prod. Sci.* 86, 35–45. doi: 10.1016/s0301-6226(03)00151-9
- Schaeffer, L. R., and Dekkers, J. C. M. (1994). "Random regressions in animal models for test-day production in dairy cattle," in *World Congress of Genetics Applied Livestock Production* (Vancouver: WCGALP), 443–446.
- Schaeffer, L. R., Jamrozik, J., Kistemaker, G. J., and Van Doormaal, B. J. (2000). Experience with a test-day model. *J. Dairy Sci.* 83, 1135–1144. doi: 10.3168/jds.s0022-0302(00)74979-4
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136
- Strabel, T., Szyda, J., Ptak, E., and Jamrozik, J. (2005). Comparison of random regression test-day models for polish Black and White cattle. *J. Dairy Sci.* 88, 3688–3699. doi: 10.3168/jds.s0022-0302(05)73055-1
- Swalve, H. (1995). The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. *J. Dairy Sci.* 78, 929–938. doi: 10.3168/jds.s0022-0302(95)76708-x
- Van Der Werf, J. H. J., Goddard, M. E., and Meyer, K. (1998). The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81, 3300–3308. doi: 10.3168/jds.s0022-0302(98)75895-3
- White, I. M. S., Thompson, R., and Brotherstone, S. (1999). Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82, 632–638. doi: 10.3168/jds.s0022-0302(99)75277-x
- 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. doi: 10.1016/0301-6226(87)90003-0
- Wood, P. (1967). Algebraic model of the lactation curve in cattle. *Nature* 216:164. doi: 10.1038/216164a0
- Yearbook, C. D. (2016). *China Dairy Yearbook*. Beijing: Dairy Association of China.

Conflict of Interest: XD was employed by the company Linqing Rutai Animal Husbandry Co., Ltd, Liaocheng, China. PB and GX were employed by the company Shandong OX Livestock Breeding Industry Co., Ltd, Jinan, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Li, Gao, Madsen, Li, Liu, Bao, Xue, Gao, Di and Su. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.