The Journal of Animal & Plant Sciences, 24(4): 2014, Page: 1045-1049
ISSN: 1018-7081

COMPARISON OF WOOD AND CUBIC SPLINEMODELS FOR THE FIRST LACTATION CURVE OF JERSEY COWS

S. Cankaya[1,] M. Sahin[2] and S. H. Abaci[3]

[1]Department of Biostatistics, Faculty of Medicine, University of Ordu, Ordu, Turkey
[2]Department of AnimalScience, Faculty of Agriculture, Kahramanmaras Sutcu Imam University, Kahramanmaras, Turkey
[3]Department of AnimalScience, Faculty of Agriculture, University of Ondokuz Mayis, Samsun, Turkey
Corresponding Author E-mail: sonercankaya@gmail.com

ABSTRACT

The aim of this study was to compare two different lactation curve models (Woodand cubic spline regression model in two knots: CSR1 and CSR2), and to find the best model that provided a good description of the first lactation curve of Jersey cattle herd.A total of 5304test-day milk yield records from monthly recording of 533the first lactation between 1984 and 2008 in Karakoy Agricultural State Farm in Samsun (Turkey). To compare the models, mean square error, coefficient of determination, adjusted coefficient of determination and Durbin Watson autocorrelation values were used. The results showed that CSR2 model which has minimum mean square error (1.910), maximum R^2 (0.910) and adjusted R^2 value (0.799) performed the best fit to the data and allowed a suitable description of the lactation curve. It was understoodthat the CSR2 model provided accurate estimates of milk yield for first lactation. Consequently, the usage of CSR2 model would provide some useful information on management decisions and genetic evaluation of the Jersey cows for milk production.

Keywords: Cubic spline, Lactation Curve, Jersey, Knot

INTRODUCTION

Milk production is one of the main incomes for dairy cattle farms, and therefore milk yield records have great deal of importance for the dairy herds (Cankaya *et al.* 2011). Lactation curve is a graphic presentation of variations in milk production throughout the lactation period (Anderson and Jenssen, 2011). Inbreeding programs shape of lactation curves and related parameters, for a successful selection widely used sources of information that the development of decision and strategies (Ozyurt and Ozkan, 2009). Therefore, lactation curves modeling with appropriate and accurate equations are of great importance.Various methods have been proposed for analyzing lactation period such as Wood, Ali Schaeffer, Wilmink, Random regression, etc (Macciotta *et al.* (2005); Takma and Akbas (2009); Cankaya *et al.*(2011). Moreover,splines are a useful function-type used in regression when the relationship between a response and a set of covariates is not known a priori (Walker *et al.* 2010) for describing the lactation curve of dairy cows.So,cubic spline regression models have been used to model lactation period with test day milk yield instead of the others(White and Brotherstone (1997); Jamrozik *et al.*(2009); Sahin and Efe (2010); Geha *et al.(*2011); Koncagul and Yazgan (2011)), because cubic spline regressions have high compliance excellence in modeling of lactation curves. In addition, more flexible curves can be obtained with the increase in the number of knots (Sahin and Efe, 2010). Accordingly, the aim of this study is to compare the Wood and cubic spline regression model in two knots (CSR1 and CSR2)used in modeling of lactation curves,and to find the best model that provided a good description of the first lactation for Jersey herds.

MATERIALS AND METHODS

Data collection: Data were 5304 test day records, which were 1st lactation official milk yield records, on 533 Jersey cows calving from 1984 to 2008 in herd under pasture-based dry seasonal production system in Karakoy Agricultural State Farm in Samsun (Turkey). Also, each data set was composed of the test days and the total amount of milk at the morning and evening milking of the test days. Lactation length varied from 240 to 308 days.

Methods: The first lactation measurements are analyzed by two methods (Wood and cubic spline regression model in two knots: CSR1 and CSR2) for description of the lactation curve using SAS statistical package program.Here, the methods used for estimation parametersand comparison criteriain this studycould be introduced as follows.

The Wood Model: The mathematical model to describe the lactation curve of dairy cows proposed Wood (1976) is one of the most popular models (Sherchand *et al.* 1995). Wood's equation is

$$Y_t = at^b \exp(-ct)$$

Where;

Y_t is milk yield in lactation day t. a is a scaling factor representing yield at the beginning of lactation, b is the inclining slope parameter up to peak yield, and c is the declining slope parameter (Silvestre *et al.*(2006); Cankaya*et al.* (2011)).

CubicSplineRegression: Splines are usually defined as piecewise polynomials of degree n with function values. The join points are called knots. Cubic splines are smooth at knots (function, first and second derivatives agree). Cubic spline regression, the knot points, usually from the inner or outer convex near the start or end points are selected. On the other hand, cubic spline regression model fit, rather than from the position of theknots, the number ofknotseffect. The number of knots increases, the number of parts increases. Therefore, increasing the number of knots usually increases the fit of the spline function to the data.

Cubic spline regression, without any requirement of an endpoint, the number of parameters needed, $\beta_0$except, " $k+3$ "is the number (Stone and Koo, 1985). In this case, one (a) and two (a,b) knots cubic spline functions regression occurs as follows.

$$Y(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-a)^3$$

and

$$Y(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t-a)^3 + \beta_5 (t-b)^3$$

A cubic spline regression model

$$y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \sum_{i=1}^{k} k(t-t_i)^3$$

Where, t refers to days in milk at the test;β_0, β_1 and $\beta_3$are parameters to be estimated, and k is the number of knots in the splines

Comparison Criteria for Models: Goodness of the fit of models was evaluated according to following criteria (Vargas *et al.* 2000):

Adjusted multiple coefficient of determination $(\mathbf{R^2_{adj}})$

$$R^2_{adj} = 1 - \left[\frac{(n-1)}{(n-p)}\right] * (1 - R^2)$$

Where;

R^2: Multiple coefficient of determination, $[R^2 = 1 - (RSS/TSS)]$

RSS: residual sum of squares,

TSS: total sum of squares,

n: number of observations,

p: number of parameters in the model.

The R^2 value is an indicator measuring the proportion of total variation about the mean $\overline{Y}$ explained by the lactation curve model. The coefficient of determination lies always between 0 and 1, and the fit of a model is satisfactory if R^2 is close to unity.

The MSE procedure gives us an asymptotic estimate of σ^2 with which we can choose a model.

MeanSquareError (MSE);

$$MSE = RSS/(n-p)$$

Where;

RSS: residualsum of squares

n: number of observations,

p: number of parameters in the model

Durbin-Watson autocorrelation test (DW): The Durbin–Watson statistic is a test statistic used to detect the presence of autocorrelation (a relationship between values separated from each other by a given time lag) in the residuals (prediction errors) from a regression analysis. Ife_t is the residual associated with the observation at time *t*, thenthe test statistic is,

$$DW = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

Wheren is thenumber of observations. Since DW is approximatelyequalto 2(1 − r), where r is the sample auto correlation of the residuals (Gujarati, 2003), DW = 2 indicates no auto correlation. Thevalue of d always lies between 0 and 4. IftheDurbin–Watson statistic is substantiallylessthan 2, there is evidence of positiveserialcorrelation. As a roughrule of thumb, ifDurbin–Watson is lessthan1.0, theremay be causefor alarm. Small values of d indicatesuccessive error terms are, on average, close in value tone another, orpositivelycorrelated. IfDW> 2, successive error terms are, on average, much different in value from one another, i.e., negatively correlated. Inregressions, this can imply an underestimation of the level of statistical significance (Wikipedia, 2013).

Mann Whitney U statistic: This is a method for the comparison of two independent random samples. The Mann Whitney U statistic is defined as:

$$U = n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum_{i=n_1+1}^{n_2} R_i$$

Where samples of size n_1 and n_2 are pooled and R_i are the ranks.

RESULTS

The focus of this study was to compare cubic spline interpolation with traditionally used the Wood model. For this aim, the applicability of cubic spline regression in two knots (CSR1 and CSR2) was investigated for modeling of individual first lactation curves of Jersey cattle herd, and to compare with Wood model.

The mean square error, coefficient of determination, adjusted coefficient of determination and DW autocorrelation values with standard errors for cubic spline regression (CSR1 and CSR2) and Wood models were given in Table 1.

Table 1.The results of the comparison criteria's for the models

Models	$\bar{X} \pm S_{\bar{x}}$			
	MSE	R^2	R^2_{adj}	DW
Wood	2.653 ± 0.100	0.776 ± 0.006	0.711± 0.009	2.209 ± 0.023
CSR1	2.054 ± 0.084	0.878 ± 0.004	0.781 ± 0.007	2.768 ± 0.017
CSR2	1.910 ± 0.083	0.910 ± 0.003	0.799 ± 0.007	3.093 ± 0.005

MSE: Mean Square Error., R^2:Coefficient of Determination, R^2_{adj}:Adjusted Coefficient of Determination, DW: Durbin-Watson Statistic (The critical value: 3.121)

When the valueswere examined in Table 1, the MSE value of Wood models (2.653 ± 0.100) higher (P<0.01) than the values of cubic spline regression models (2.054 ± 0.084; 1.910 ± 0.083). On the other hand the coefficient of determination and adjusted coefficient of determination values of Wood models (0.776 ± 0.006, 0.711 ± 0.009) lower than one (0.878 ± 0.004, 0.781 ± 0.007) and two (0.910 ± 0.003, 0.799 ± 0.007) knots cubic spline regression values (P<0.01). Also, DW auto correlation values are examined, there was not seen autocorrelation (DW<3.121) in all three models.

The mean square error, coefficient of determination and the adjusted coefficient of determination of models and the results of comparison Mann-Whitney U tests given in Table 2. MSE, R^2 and R^2_{adj} values are not normally distributed. So, we used the Mann Whitney U tests for comparison of models with MSE, R^2andR^2_{adj}.

Table 2.The results of Mann-Whitney U test for comparison of the models with MSE, R^2 and R^2_{adj}

Criteria	Model Comparison	$\bar{X} \pm S_{\bar{x}}$	Mann-Whitney U	
			Z	Sig.
MSE	Wood- CSR1	2.653 ± 0.100 - 2.054 ± 0.084	-5.198	.000
	Wood- CSR2	2.653 ± 0.100 - 1.910 ± 0.083	-7.210	.000
	CSR2- CSR1	1.910 ± 0.083 - 2.054 ± 0.084	-2.274	.023
R^2	Wood- CSR1	0.776 ± 0.006 - 0.878 ± 0.004	-11.48	.000
	Wood- CSR2	0.776 ± 0.006 - 0.910 ± 0.003	-16.27	.000
	CSR2- CSR1	0.910 ± 0.003 - 0.878 ± 0.004	-6.284	.000
R^2_{adj}	Wood- CSR1	0.711 ± 0.009 - 0.781 ± 0.007	-5.549	.000
	Wood- CSR2	0.711 ± 0.009 - 0.799 ± 0.007	-7.728	.000
	CSR2- CSR1	0.799 ± 0.007 - 0.781 ± 0.007	-2.462	.014

Z; z-statistic, Sig.; level of importance.

When the cubic spline regressions is assessed in itself, the model of two knots' mean square error (P<0.05), coefficient of determination (P<0.01) and adjusted coefficient of determination (P<0.05) have been better results than the model of one knot(Table 2). However, increasing the number of knots in case of cubic spline regression, Durbin Watson autocorrelation values for the studied data set is seen approaching the limit value.The coefficient of Wood and Cubic Spline Regression models (CRS1 and CRS2) given in Table 3.

Table 3. Coefficientsfor Wood and Cubic Spline Regression Models.

		$a \pm S_{\bar{x}}$		$b \pm S_{\bar{x}}$		$c \pm S_{\bar{x}}$	
Wood		4,634411 ± 1,712944		0,425200 ± 0,123378		0,006314 ± 0,001116	
Cubic		β_0	β_1	β_2	β_3	β_4	β_5
CRS1	β	6,529482	0,298497	-0,003650	0,000013	-0,000015	-
	$S_{\bar{x}}$	3,799778	0,171061	0,002224	0,000009	0,000009	-
CRS2	β	5,931220	0,331314	-0,004156	0,000016	-0,000016	0,0000012
	$S_{\bar{x}}$	4,641893	0,230084	0,003298	0,000014	0,000010	0,000006

Cubic spline regressions (CSR1 and CSR2), milk yield of the Wood model and individual lactation curves generated using the control days given in Figure 1.Increase in the number of knots, in point spread, it is causes to obtain a more flexible curve and can express the better of distribution shown in Figure 1.

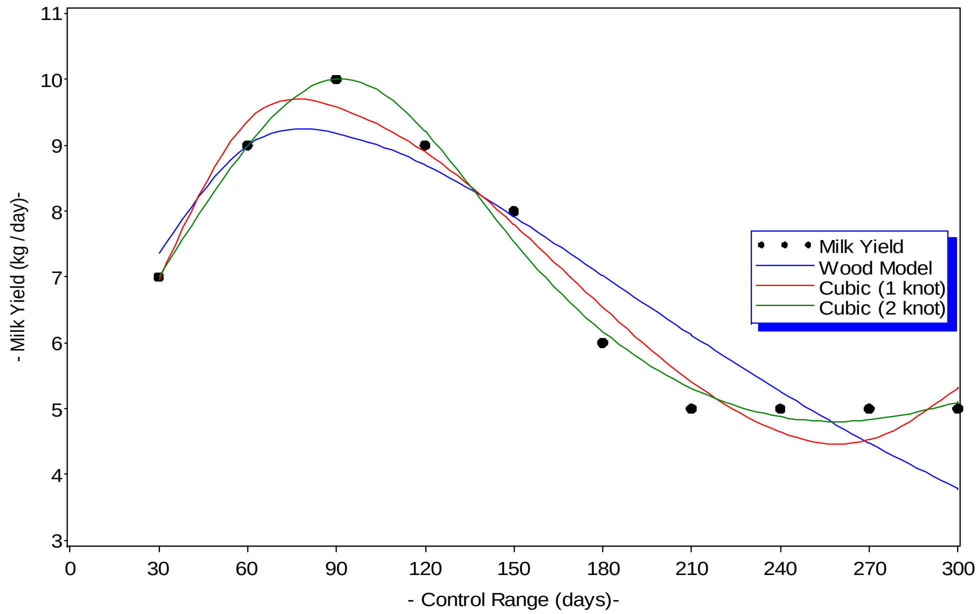

Figure 1.Individual Lactation Curves

DISCUSSION

According to model comparison criteria's used the study, cubic spline regression models have the best results in modeling of individual lactation curves. Moreover, within these models, cubic spline regression in two knots can be said that the best model which have minimum MSE and maximum R^2 and R^2_{adj}.

The findings of this study is consistent with the comparative studies (Druet *et al.* (2003); Silvestre *et al.* (2006)) used Wood models and cubic spline regression models. The findings is also consistent with the studies (White *et al.* (1999); Sahin and Efe (2010); Nicolo *et al.* (2010); Koncagul and Yazgan (2011); Geha *et al.*(2011)) used cubic spline regressionthatmodeling of lactation curves of Holstein Friesian dairy cattle.Therefore, it can besaid that cubic spline regression models' that having with appropriate position and number of knotscan be used the modeling of individual lactation curves in Jersey Cows.

Consequently, the usage of CSR2 model estimated as the best model for a good description of Jersey cows' the first lactation would provide some useful information on management decisions and genetic evaluation of the Jersey cows for milk production.

Acknowledgements: The authors wish to thank the staff of Karakoy State Farm for technical assistance.

REFERENCES

Anderson, F., and M. Jenssen, (2011) [Internet]. Use of the lactation curve in the production management of Norwegian dairy cattle [cited 2011 December] Available from: http:// www.veths.no/en/ Home/ News/News-stories/ Use-of-the-lactation-curve-in-the-production-management-of-Norwegian-dairy-cattle

Cankaya, S., A. Unalan, andE. Soydan, (2011). Selection of a mathematical model to describe the lactation curves of Jersey cattle. Arch Tierz. 54 (1): 27-35.

Druet, T., D. B. Jaffrezic, and V. Ducroco (2003). Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. J. Dairy Sci. 86 (7): 2480-2490.

Geha, M.J., J.F. Keown and L.D.V. Vleck (2011). Cubic spline interpolation to estimate effects of inbreeding on milk yield in first lactation Holstein cows. Genetics and Molecular Biology. 34 (3): 443-450.

Gujarati, D.N. (2003). Basic econometrics.4th Ed. McGraw-Hill; Boston.

Jamrozik, J., J. Bohmanova and L.R. Schaeffer (2009). Selection of locations of knots for linear splines in random regression test day models. J. Anim. Breed. Genet. 127 (2): 87-92.

Koncagul, S., and K. Yazgan (2011). Comparison of spline and Legendre polynomial functions with conventional methods for describing lactation curves of Holstein dairy cows. J. Anim. Production. 52(1): 17-23.

Macciotta, N.P.P., D. Vicario, and A.Cappio-Borlino (2005). Detection of different shapes of lactation curve for milk yield in dairy cattle by empirical mathematical models. J. Dairy Sci. 88 (3): 1178-1191.

Nicolo, P.P.M., M. Filippo, D. Corrado, R.S. Larry (2010). Comparison of parametric, orthogonal, and spline functions to model individual lactation curves for milk yield in Canadian Holsteins. Ital. J. Anim. Sci. 9 (87): 460-464.

Ozyurt, A., and M. Ozkan (2009). The lactation shape and effecting factors on lactation curves of Holstein cows bred in middle Anatolia.J. Anim. Prod. 50 (1): 31-37.

Sahin, M. andE. Efe (2010).Use of cubic spline regressions in modeling dairy cattle lactation curves.KSU. J. Natural Sci. 13 (2): 17-22.

SAS Institute Inc., 1999. SAS Online Doc®, Version 8, Cary, NC: SAS Institute Inc. [cited 2013 Feb 12] Available from: http://v8doc.sas.com/sashtml/stat/chap65/sect43.htm

Sherchand, L., R.W. Mcnew, D.W. Kellogg, and Z.B. Johnson (1995). Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. J. Dairy Sci. 78 (11): 2507-2513.

Silvestre, A.M., T. Petim-Batista, and Colaço (2006). The accuracy of seven mathematical functions in modeling dairy cattle lactation curves based on test-day records from varying sample schemes. J. Dairy Sci. 89 (5): 1813-1821.

Stone, C.K., andC.Y. Koo (1985).Additive splines in statistics, In Proceedings of the Statistical Computing Section. American Statistical Association, 45-48.

Takma, C., and Y. Akbas (2009). Comparison of fitting performance of random regression models to test day milk yields in Holstein Friesians. Kafkas Univ. Vet. Fak. Derg. 15 (2): 261-266.

Vargas, B., W.J. Koops, M. Herrero, and J.A.M. Van Arendonk (2000). Modeling extended lactations of dairy cows. J. Dairy Sci, 83 (6): 1371-1380.

Walker, C.G., M.J. O'Sullivan, Y. Zhu, andS. Hsu (2010).Regression spline fitting with applications. Proceedings of the 45th Annual Conference of the ORSNZ. 271-.280.

White, I., and S. Brotherstone (1997). Modelling lactation curves with cubic splines. Interbull Bulletin, (16), 80.

White, I.M.S., R. Thompson, andS. Brotherstone (1999). Genetic and environmental smoothing of lactation curves with cubic splines. J. Dairy Sci. 82 (3): 632-638.

Wikipedia (2013). Durbin–Watson statistic; [cited 2013 April 20] Available from: http://en.wikipedia.org/wiki/Durbin%E2%80%93Watson_statistic

Wood, P.D.P., (1976). Algebratic models of the lactation curves for milk, fat and protein production with estimates of seasonal variation. Anim. Prod. 22 (1): 35-40.