Irish Journal of Agricultural and Food Research **44:** 000–000, 2005

Empirical algebraic modelling of lactation curves using Irish data

N. Quinn[1,2,†], L. Killen[1] and F. Buckley[2]
[1] *School of Computing, Dublin City University, Dublin 9*
[2] *Dairy Production Research Centre, Teagasc, Moorepark, Fermoy, Co. Cork*

The purpose of this study was to find a well-fitting, robust, single equation model to describe the shape of lactation curves for Irish dairy cows. The suitability of a number of algebraic models that depict lactation curves was examined, using Irish test day data. The analysis was carried out on a total of 14,956 lactation records from commercial and experimental herds and included both autumn and spring calving animals. 'Goodness of fit' and adherence of the various models to the assumptions of regression analysis were examined. Multicollinearity posed a severe problem in the application of the best-fit model but omitting one of the variables from the estimation procedure reduced this effect. The modified model, referred to as the Ali-B model, is a single equation model that can be easily updated and incorporated into computer code. This is in contrast with the Standard Lactation Curve (SLAC) method, a method of interpolation, which is currently adopted by the Irish industry. The effects of seasonal factors on milk production were estimated and added to the Ali-B model to create a production profile for cows calving in specific months. The Ali-B model provides an acceptable level of accuracy in representing the shape of the lactation curve for Irish dairy cows, and can be easily modified for different environmental scenarios.

Keywords: Analysis of residuals; dairy cows; lactation curves; models

Introduction

Empirical algebraic modelling of lactation curves offers a summary of longitudinal milk yield patterns from which cumulative lactation curves can be estimated or by which total lactation milk yields may be predicted

†Corresponding author: Tel: + 353 (0)1 700 8449; Fax: +353 (0)1 700 5442; e-mail: nquinn@computing.dcu.ie

from incomplete data. Appropriate models provide useful information for breeding and management decisions at both industry and farm level, and also for the comparison of alternative production strategies in bio-economic modelling. To ensure accurate decisions pertinent to individual animals or herds it is essential that cumulative yield is predicted with minimum error and from relatively few test dates, the latter reducing the cost and inconvenience of milk recording. From the bio-economist's viewpoint, the lactation curve model must accurately depict what is expected at farm level so that a cost can be associated with each cow.

Many authors have contributed to the evolution of lactation curve modelling by using empirical regression models, test-day models (Keown and Van Vleck, 1972; Schaeffer and Burnside, 1976), multiphasic models (Grossman and Koops, 1988; Sherchand *et al*., 1995), Bayesian analysis (Jones, 1997; Jamrozik, Ginola and Schaeffer, 2001) and autoregressive procedures (Carvalheira *et al*., 2001; Vasconcelos *et al*., 2002). This study will focus on empirical regression models because they have been found to perform well statistically over a wide variety of datasets; they are often biologically interpretable and they are easy to apply, which is of great benefit to scientists and economists. The models investigated in this study are shown in Table 1. The model of Wood (1967) is the basis for studies involving empirical equations (Pérochon, Coulon and Lescourret, 1996) and this was the root of this investigation. Wood's model:

$$Y_n = an^b e^{-cn}, \tag{1}$$

where Y_n is the yield in week n, uses the method of least squares to obtain estimates for three parameters in the incomplete gamma function: a is a scaling factor associated with the average yield, b is related to pre-peak curvature and c is related to post-peak curvature. Many alternative models (Keown and Van Vleck, 1972; Yadav *et al.*, 1977; Cobby and Le Du, 1978; Ali and Schaeffer, 1987; Wilmink, 1987) have been proposed as a consequence of the lack of fit of Wood's

Table 1. Selection of models investigated

Year	Researcher	Model*
1923	Brody *et al*.	$Y_n = ae^{-bn}$
1924	Brody *et al*.	$Y_n = ae^{-bn} - ae^{-cn}$
1950	Sikka	$Y_n = ae^{(bn-cn^2)}$
1967	Wood	$Y_n = an^b e^{-cn}$
1967	Wood	$\ln(Y_n) = \ln(a) + b\ln(n) - cn$
1971	Dave	$Y_n = a + bn - cn^2$
1977	Cobby & Le Du	$Y_n = a - bn - ae^{-cn}$
1977	Yadav	$Y_n = \frac{n}{a+bn+cn^2}$
1979	Madalena *et al*.	$Y_n = a - bn$
1979	Molina & Boschini	$Y_n = a - b\|n - c\|$
1982	Singh & Gopal	$Y_n = a - bn + c\ln(n)$
1987	Ali & Schaeffer	$Y_n = a + b\gamma + c\gamma^2 + d\omega + e\omega^2 + f^{**}$
1987	Wilmink	$Y_n = a + be^{-kn} + cn$
1995	Guo & Swalve	$Y_n = a + b\sqrt{n} + c\ln(n)$

* Y_n is the yield in lactation week n

** $\gamma = \frac{7n}{305}, \omega = \ln\left(\frac{305}{7n}\right)$

model under certain circumstances. Models such as those proposed by Yadav *et al.* (1977) in India and Wilmink (1987) in Canada were developed for this reason and in subtropical and tropical climates, Kellogg, Urquhart and Ortega (1977) and Shanks *et al.* (1981) found Wood's model very poor in fitting data. Examination of Cobby and Le Du's (1978) equation reveals that it is a variation of Wood's model, where e^{-cn} is replaced by where $(1 - e^{-qn})$ and q are parameters, and n is lactation week. The model of Wilmink (1987) is a variation of the model of Cobby and Le Du and therefore also a derivative of the model of Wood (1967). Killen and Keane (1978) tested Wood's model under Irish conditions and obtained R^2 values of the order of 0.7; they examined the shape of lactation curves for milk, fat and protein production in the Irish context.

In Ireland presently, the SLAC (Standard Lactation Curve) method of Olori and Galesloot (1999) is used for predicting milk yield. This is a method of interpolation incorporating 2,160 lactation curves, accounting for variation in the effects of season, calving age and level of production. While it is acknowledged that having a library of equations, from which the most appropriate one is chosen, will almost inevitably give accurate predictions, a single-equation model has many advantages. For example, it is simple to incorporate it into computer programs and is easily updated and can be easily re-examined in the light of new data. A single-equation model would also be considered more appropriate for use by bio-economists who need to constantly update and re-create the parameters for different scenarios.

Advances in management technology, cow production potential and procedures to evaluate lactation curve models have resulted in a renewed interest in re-examining empirical regression procedures under Irish conditions. Alternatives to Wood's model, such as those proposed by Ali and Schaeffer (1987), Wilmink (1987), and Guo and Swalve (1995) are considered worthy of investigation, because they have been proven to have a better fit than Wood's model in the respective studies. The objectives of this study were to compare the goodness-of-fit of numerous empirical models including the models of Wood (1967), Cobby and Le Du (1978), Ali and Schaeffer (1987), Wilmink (1987) and Guo and Swalve (1995), to analyse the residuals of each model in an attempt to find a well-fitting, robust, single equation model, for weekly milk yields, by examining the assumptions of regression, to determine the reliability of the different models in estimating total milk yield which is tested under present day Irish production conditions and also to provide a seasonality production pattern table for use by bio-economists.

The assumptions of a regression model that need to be examined to assess the "appropriateness" of a model are (1) independence of the error terms (no autocorrelation), (2) independence of the explanatory variables (absence of multicollinearity), (3) constancy of the variance of error terms (homoscedasticity) and (4) normality of the distribution of error terms. It has been shown previously (Maddala, 1992) that the assumptions required for fitting models using regression analysis are not always satisfied. For the non-linear equations represented by the model, $Y_u = f(\zeta_u, \theta) + \varepsilon_u$ with $\varepsilon \sim N(0, \sigma^2)$ where Y_u is the dependent or response variable, ζ_u is an independent or predictor variable and θ are the parameters, the true residuals, $\varepsilon_1, \ldots \varepsilon_u$, are assumed to be normally distributed (N) with zero mean and constant variance of σ^2.

Cunningham (1972) found that lactation number, calving month and herd had significant effects on milk yield in Ireland

and that the effects of these variables should be included. Milk yield is also affected by other environmental factors, such as weather, and feeding regime. In particular there is a stimulus to milk production resulting from high availability and digestibility of grass in spring and a depressing effect due to the use of conserved forage in the winter (Cunningham, 1972; Killen and Keane, 1978). It was necessary that this be examined so that a seasonal production pattern table could be created for use by bio-economists, which would account for the variation caused in milk yield due to production month.

Materials and Methods

Data

Two sets of data were available for this study; Dataset 1 comprised 14,198 lactations, with monthly test day yields recorded during 1999 and 2000 from 79 commercial spring-calving dairy herds. Within this dataset 4,336 cows had repeated lactations across the 2 years. All cows were spring/summer calving (SSC), defined as calving from January to June. Test-day records for individual cows were obtained from the Irish Dairy Records Cooperative (IDRC) and milk recording was carried out at 4-week intervals. Dataset 2 comprised lactation records from six research herds (Teagasc) recorded over the period 1995–2001. This dataset included 1,888 lactations from 872 individual cows, of which 1,408 lactations were SSC and the remainder were autumn/winter calving (AWC), defined as calving from July to December. In these herds, individual milk yield was recorded weekly. In Dataset 1, concentrate supplementation per cow per lactation averaged 745 kg, with a range from 335 kg to 1,305 kg per cow per lactation, for individual herds. In Dataset 2, the average level of concentrate supplementation was 500 kg, with a range of 300–700 kg, for the SSC and approximately 1,500 kg for the AWC animals. The majority of lactations in the study were from Holstein-Friesian cows, with the exception of 142 lactations from pure Normande cows and 147 lactations from pure Montebeliarde cows. These were present on one of the farms, attached to Moorepark, in Dataset 2.

Data collation and editing

All herds in the study were incorporated into the Dairy Management Information System (DairyMIS) operated by Moorepark Research Centre (Crosse, 1986). DairyMIS is a recorder-based computerised system for collecting detailed data on stock numbers, farm inputs, production, and reproduction information on a monthly basis. The lactation number of each cow was also obtained from the IDRC files and calving date records were captured through the DairyMIS system. The calving date and drying-off date (also obtained from the IDRC files) were used to validate lactation number and test-day records for a given lactation. Lactation number ranged from 1 to 16 across the datasets. For the purposes of this study, lactation number was categorised as lactation 1, lactation 2 and lactation 3 or greater (Cunningham, 1972; Killen and Keane, 1978; Lidauer *et al*., 2000; Dechow *et al*., 2004); 4,093 records were for cows in their first lactation, 3,722 for cows in their second lactation and 8,112 in their third or greater lactations (see Table 2). The average yield per lactation, as shown in Table 2 also, was calculated only for Dataset 2 because there were only part-lactation records available for Dataset 1. It can be seen that, for example, a cow in its first lactation produces on average 5438 kg of milk while a cow in their third or greater lactation produces approximately 6454 kg of milk over a lactation. Lactations with fewer than five recordings

Table 2. Number of animals and average total yields per lactation

		Lactation number		
		1	2	3+
Number of animals	Experimental	585	484	660
	Commercial	3508	3238	7452
Average yield (kg)		5437.8	6324.6	6453.7
Standard deviation of yield (kg)		1024.8	1276.0	957.7

were deleted from Dataset 1 and lactations of less than 25 weeks duration were removed from Dataset 2. After edits, Datasets 1 and 2 consisted of 13,229 and 1,727 lactations respectively.

Models and Statistical Analysis

A preliminary examination was carried out on the models of Brody *et al.* (1923, 1924), Sikka (1950), Wood (1967), Dave (1971), Cobby and Le Du (1978), Yadav *et al.* (1977), Madalena *et al.* (1979), Singh and Gopal (1982), Ali and Schaeffer (1987), Wilmink (1987) and Guo and Swalve (1995). These models were fitted to the pooled data using linear and nonlinear regression and then the effects of lactation number, calving month and herd were removed from the parameters. This was done by calculating the mean parameters for each herd, and within each herd the mean parameters for each calving month, and within each calving month finding the mean parameters for each lactation number; these findings were used to give the mean parameters for each model. Those models that were found to be poor in fitting the Irish data (i.e. $R^2 < 0.60$) were eliminated from further consideration. The value of a non-linear equation was found by calculating:

$$1 - SSE/CSS \qquad (2)$$

where *SSE* is the error sum of squares and *CSS* is the corrected total sum of squares for the dependent variable. The Mean Square Prediction Error (MSPE) value was also used as a measure of goodness-of-fit (Kvanli, Guynes and Pavur, 1986) using the following formula:-

$$MSPE = \frac{\sum_{t=1}^{n} e_t^2}{n} \qquad (3)$$

where e_t is the residual for observation t and n is the number of predicted values obtained. This was calculated for each herd before calculating the overall mean MSPE value for each model. After the preliminary examination, the better models, those of Wood (non-linear form), Wilmink, Ali and Schaeffer, and Guo and Swalve, were considered to have acceptable MSPE values (MSPE values <580) and further analysis was performed on these models.

Wood's linear model was examined in three forms, as discussed by Cobby and Le Du (1978), initially in its linear form without statistical weights, later by weighting the logarithm of the milk yield (i.e. the dependent variable) proportionally to the square of the milk yield and finally in its nonlinear form. When examining Wilmink's model, an analysis of variance was carried out on parameter k similar to Olori *et al.* (1999), and Brotherstone, White and Meyer (2000). It was found that k was constant, with a value of 0.10, over lactation number, calving month and herd effect for the Irish data used in this study. The residuals of these models together with those of Ali and Schaeffer (1987) and Guo and Swalve (1995) were then analysed in detail.

The Durbin-Watson statistic was calculated for all the models to test for the existence of autocorrelation between the residuals (Mendoza, 1999). A consequence of autocorrelation is that the degrees of freedom are over-estimated, leading to unrealistic t statistic values. The decision rules for autocorrelation used in this paper are those outlined by Mendoza (1999). Initially, first order autocorrelation was examined and if this proved to be inconclusive then higher-order autocorrelations were tested. A condition index was calculated to test for the presence of multicollinearity. If the values ranged from 30 to 100 it indicated that moderate to strong multicollinearity existed (Belsley, Kuh and Welsch, 1980). When multicollinearity exists two problems may occur: the computation of the parameter estimates may be slow and nonconvergent, and the parameter estimates may have inflated variances (Belsley *et al.*, 1980). To reduce multicollinearity at least one of the variables needs to be removed. The test for violation of homoskedasticity is White's test (Sen and Srivastava, 1990) and this was calculated for each individual lactation, a mean value being computed after accounting for calving month, lactation number and herd effect. Additional tests included the Kolmogorov-Smirnov statistic (a test for normality of the distribution of the residuals), skewness and kurtosis. A t-test was performed on the kurtosis and skewness values to test if they were significantly different from zero; the kurtosis value of a normal distribution being zero.

As the total milk yield is only known for the cows in Dataset 2, the models of Wood (1967), Wilmink (1987), Ali and Schaeffer (1987), and Guo and Swalve (1995) were used to estimate the total milk yield for each cow only in this dataset. This was performed by first determining the parameter values for each lactation number, which were then used to predict the mean weekly yields for each cow; the time variable in all the models is measured in weeks. The weekly yields were then cumulated to give the estimated total milk yield for each cow for each model. The mean of the estimated total milk yield and the actual total milk yield were then calculated by weighting them according to the number of animals in each lactation number category.

The deviations resulting from comparing the best model with the actual data, from both datasets, were cumulated for each month of the year, and the mean of the deviations for each month was computed. This was used to estimate the effect of some environmental factors on a seasonal basis. The seasonal effects were then averaged over several seasons from 1995 to 2002 to take account of possible year-to-year environmental variations. To compute the percentage of the total lactation yield in each month throughout lactation, the yield predicted by the chosen model was adjusted by these seasonal effects.

Results

Goodness of fit

The preliminary examination resulted in the models of Brody *et al.* (1923, 1924), Sikka (1950), Dave (1971), Yadav *et al.* (1977), Madalena *et al.* (1979), Molina and Boschini (1979) and Singh and Gopal (1982) being eliminated because they gave rise to very high MSPE values (greater than 580), indicating that these models fitted to data poorly. In fact Brody *et al.*'s (1924) equation failed to converge. The goodness-of-fit statistics of the expected curves for weekly milk yield, for the better-fitting models and the model of Wood

Table 3. Goodness of fit statistics of expected curves for weekly milk yield

Model	MSPE
Wood (Linear form)	583.99
Wood (Weighted linear form)	624.83
Wood (Non-linear form)	562.16
Wilmink	603.72
Ali & Schaeffer	501.79
Ali-B	520.93
Guo & Swalve	556.57

(1967), are presented in Table 3; MSPE values ranged from 624.83 to 501.79. Ali and Schaeffer's model gave the best fit with a MSPE value of 501.79, while Wood's model, in weighted linear form, gave the poorest fit (MSPE value of 624.83).

Analysis of residuals

The analyses of residuals were performed on the best four models and these included the models of Wood (1967) (nonlinear form), Wilmink (1987), Ali and Schaeffer (1987) and Guo and Swalve (1995). The Durbin-Watson statistic, d, was found to be between d_u and $4 - d_u$ (where d_u is the upper critical bound as outlined by Mendoza (1999)), for first-order autocorrelation, for all of these models, indicating autocorrelation was not present.

Examination of the multicollinearity diagnostics revealed that in Ali and Schaeffer's model multicollinearity was severe, with a condition index value of 1075.4 (Table 4). In Guo and Swalve's model it was moderate (condition index value of 49.39), whereas in Wilmink's and Wood's models, multicollinearity was weak (condition index values of 15.32 and 25.20, respectively). Ali and Schaeffer's model was then examined, with each of the variables removed in turn (thus removing a parameter each time the model was re-estimated). It was found that the condition index could be reduced when parameter *b, c, d* or *e* was removed, but the greatest improvement occurred when parameter *b* was removed. The MSPE value for this new model (without parameter *b*) denoted Ali-B, was 520.93, which is a little higher than the MSPE value found when fitting Ali and Schaeffer's original model (Table 4). It was found that all models had a P-value for White's test of greater than 0.05, indicating that heteroskedasticity was not a problem in any of the models. The assumption of normality was found not to be a problem, with the value of the Kolmogorov-Smirnov test statistic, *D,* varying in value from 0.10 to 0.11 across the models investigated. Kurtosis varied between 0.34 and 0.78, and skewness varied between –0.04 and –0.18, again indicating the assumptions of normality were valid.

The models used to estimate the expected lactation curves for milk yield are shown in Table 5. The *b* parameter values

Table 4. Comparison of models

Test	Wood	Wilmink	Ali & Schaeffer	Ali-B	Guo & Swalve
MSPE	562.16	603.72	501.79	520.93	556.57
R^2	0.63	0.60	0.68	0.67	0.64
Autocorrelation	No 1st Order	No 1st Order	No 1st Order	No 1st Order	No 1st Order
Multicollinearity	Weak	Weak	Severe	Moderate	Moderate
(Condition Index)	(25.2)	(15.3)	(1075.4)	(55.9)	(49.4)
Heteroskedasticity	No	No	No	No	No
Normality	Normal	Normal	Normal	Normal	Normal
Kurtosis	0.60	0.78	0.53	0.34	0.41
Skewness	–0.07	–0.18	–0.07	–0.05	–0.06

Table 5. Expected curve models for milk yield

Model	Equation estimated*
Wood (Linear form)	Y_n = exp(119.87 + 0.35 ln(n) – 0.04n)
Wood (Weighted linear form)	Y_n = exp(142.02 + 0.27 ln(n) – 0.04n)
Wood (Non-linear form)	$Y_n = 143.13n^{0.32} e^{-0.04n}$
Wilmink	$Y_n = 262.37 - 102.46e^{-0.10n} - 4.53n$
Ali & Schaeffer	$Y_n = 194.42 - 99.87\gamma - 16.20\gamma^2 + 19.35\omega - 7.92\omega^{2**}$
Ali-B	$Y_n = 121.98 - 52.46\gamma^2 + 71.06\omega - 18.94\omega^{2**}$
Guo & Swalve	$Y_n = 190.18 - 71.49\sqrt{n} + 95.20\ln(n)$

*Y_n is the yield in lactation week n

** $\gamma = \frac{7n}{305}$, $\omega = \ln\left(\frac{305}{7n}\right)$

for the three forms of Wood's model were very similar at 0.35, 0.27 and 0.32 for Wood's linear, weighted linear and nonlinear forms, respectively. The values were also very similar for parameter c between the three estimation procedures (0.041, 0.035 and 0.039, respectively).

Estimation of total milk yield and seasonality of production

The average actual total milk yield for Dataset 2 was 5702 kg. Using Ali and Schaeffer's model gave an estimated total milk yield of 5937 kg, overestimating the average by almost 4% (See Table 6). The Ali-B model estimated the total milk yield to be 5795 kg, only overestimating the average by 1.6%. Wood's, Wilmink's and Guo and Swalve's models also overestimated total milk yield by 11.2%, 1.9% and 2.5%, respectively, whereas the model of Wood in both linear and weighted linear form underestimated the total milk yield.

Table 6. Comparison of estimated total yield with actual total yield

Model	Total milk yield (kg)	Percentage deviation (%)
Actual total milk yield	5702	–
Wood (Non-linear form)	6423	11.2
Wilmink	5813	1.9
Ali & Schaeffer	5937	3.9
Ali-B	5795	1.6
Guo & Swalve	5849	2.5

Table 7 shows the incremental change accounting for environmental effects on the Ali-B model, independent of stage of lactation. In May, milk production tends to be higher, an average increase on the Ali-B model of almost 9%, due to environmental effects. In December milk production is reduced due to seasonal effects by over 8% in the Ali-B model. It is evident that the summer months have a positive effect and that the winter months have a negative effect on milk production.

An example of a production profile for a cow calving on the 15th day of each month is shown in Table 8. A cow calving in mid May will produce on average approximately 5% of her total milk yield in May, 14% of her total milk yield in

Table 7. Seasonal deviations on the Ali-B model, independent of stage of lactation

Month	Milk yield (%)
January	–4.5
February	–6.9
March	1.4
April	5.9
May	8.9
June	7.0
July	8.0
August	3.3
September	–0.6
October	–2.2
November	-6.8
December	–8.6

Table 8. An example of a production pattern profile for a cow calving on the 15^{th} day of each month showing the percentage of milk yield supplied in each production month

	Calving month											
Production month	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
Jan	4.0		2.1	5.9	7.1	8.3	9.4	10.5	11.5	12.3	12.9	12.6
Feb	11.4	3.2		2.0	5.3	6.4	7.4	8.4	9.3	10.1	10.8	11.2
Mar	13.4	13.1	4.6		2.6	6.5	7.7	9.0	10.1	11.1	12.1	12.8
Apr	12.8	13.5	13.4	4.4		2.8	6.6	7.9	9.1	10.2	11.2	12.0
May	12.7	13.6	14.3	14.3	5.1		2.9	7.0	8.4	9.6	10.7	11.7
June	11.0	12.0	12.9	13.7	13.8	4.6		2.8	6.7	7.9	9.0	10.0
July	10.4	11.5	12.5	13.6	14.5	14.6	5.1		3.1	6.9	8.2	9.3
Aug	8.8	9.9	10.9	12.0	13.1	14.0	14.1	4.9		2.7	6.5	7.7
Sept	7.1	8.2	9.1	10.2	11.3	12.4	13.2	13.2	4.3		2.6	6.0
Oct	6.1	7.2	8.2	9.4	10.5	11.7	12.7	13.5	13.3	4.6		2.5
Nov	2.4	5.6	6.5	7.7	8.8	9.8	10.8	11.7	12.4	12.2	4.0	
Dec		2.5	5.6	6.7	7.9	9.0	10.0	11.0	11.9	12.4	12.2	4.2
	100	100	100	100	100	100	100	100	100	100	100	100

June, 14% of her total milk yield in July and so on. Where a cow calving in December will produce approximately 4% of her total milk yield in December, 13% in January, 11% in February, and so on.

Discussion

The main objective of this study was to examine the suitability of a number of single equation models to describe milk yield throughout a lactation. The results show that the Ali-B model is the most consistent at adhering the assumptions and predicting individual weekly and total milk yield. The model represents a considerable improvement of Wood's model, in linear form, which was used in the study of Killen and Keane (1978).

A comparison of the parameter estimates for Wood's model in Killen and Keane's study (1978) with those found in this study give an indication of changes which have occurred in dairying in the 26 years between the two studies. The parameter estimates for Wood's model and those which were found by Killen and Keane (1978) were compared and the mean values of the shape parameters have increased from 0.331 to 0.353 for parameter b and from –0.058 to –0.041 for c, meaning that the upward and downward slopes have in fact become steeper (Figure 1). In the 1978 study, peak yield was estimated to have occurred around week six of lactation (b/c), whereas in this study it was observed around week eight. In comparison with the study of Killen and Keane (1978) the average milk yield per cow, as predicted by Wood's model (in linear form), has increased more than twofold from 2364 to 5448 kg in 26 years. It must be acknowledged, however, that the two studies used quite different datasets; the data used by Killen and Keane (1978) came only from experimental herds, while the dataset in the present study, as well as being somewhat larger, includes data from commercial dairy herds.

Model Selection

The best model was selected on its MSPE value, its ability to adhere to the assumptions of regression analysis and its ability to estimate total milk yield. The MSPE

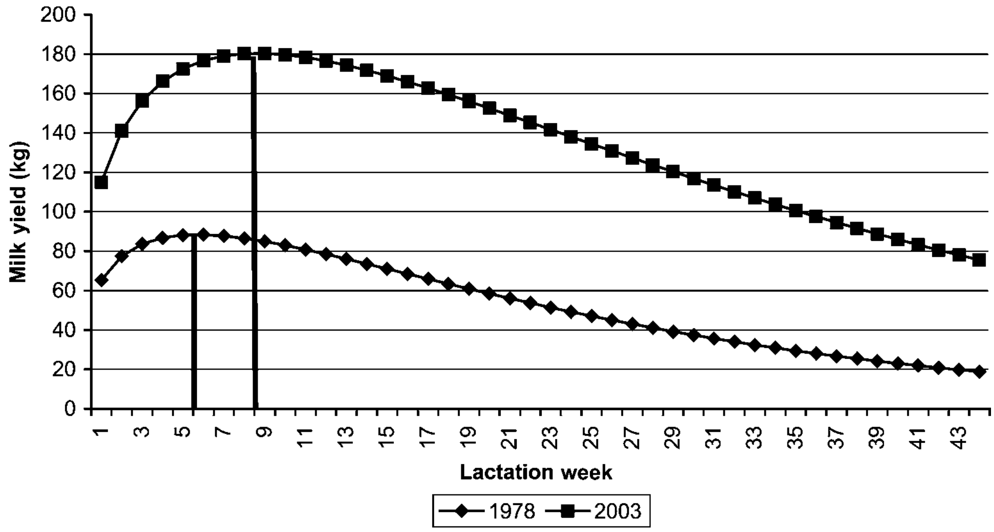

Figure 1: Comparison of lactation curves in 1978 and 2003 showing the peak week of milk yield.

values found in this study for the linear and non-linear versions of Wood's model reinforce the point made by Cobby and Le Du (1978) that non-linear regression would prove to be a more satisfactory method of fitting the model to the data.

The residuals for the better models were analysed to test adherence to the assumptions which are made when fitting the models using regression analysis. The normality assumption was not a problem although it is acknowledged that the test's ability to reject the null hypothesis increases with the sample size (SAS, 1999). It was concluded from the values of the Kolmogorov-Smirnov test statistic, kurtosis and skewness, that there was no significant deviation from normality in the distribution of the residuals.

The only assumption that was found to be a problem was that of the explanatory variables being independent in every case (multicollinearity). The condition index was extremely high for Ali and Schaeffer's model, but when parameter *b* was removed it was found that the resulting model (Ali-B) was the most satisfactory in that it satisfied all assumptions tested and multicollinearity was no longer a major issue. While there was still some multicollinearity between the explanatory variables, this is inevitable (Maddala, 1992). The Ali-B model also had a relatively good MSPE value, and it was concluded that it was necessary to sacrifice some goodness-of-fit for adherence to assumptions. Ali and Schaffer included the *b* parameter and the variable associated with it, because when it was included in the model the fit was better. However, it is clear from the analysis carried out in this study that Ali and Schaeffer's model (including the *b* parameter) does not satisfy all the assumptions of regression analysis; multicollinearity was at an unacceptable level. The MSPE value for the Ali-B model was 520.93, which is slightly higher than the MSPE value of Ali and Schaeffer's original model but the problem of multicollinearity among the independent variables inflating the standard errors was reduced.

In all, the results show that a modified Ali and Schaeffer model (Ali-B) best met the criteria of predicting the weekly milk yield of individual cows. The Ali-B model is a polynomial regression model of the following form:

$$Y_n = a + c\gamma^2 + d\omega + e\omega^2 + \varepsilon$$

where Y_n is the yield in lactation week n,

$$\gamma = \frac{7n}{305}, \omega = \ln\left(\frac{305}{7n}\right),$$

ε, is the residual and a, c, d and e are regression coefficients. The resulting equation found for this particular model is as follows:

$$Y_n = 121.98 - 52.46\gamma^2 + 71.06\omega - 18.94\omega^2$$

The Ali-B model was also found to be the best model for describing total milk yield. Accurate prediction of total milk yield will help to improve the accuracy of genetic predictions of sires and dams (Olori *et al.*, 1999; Koonawootrittriron *et al.*, 2001) and is of benefit to bio-economists (Shalloo *et al.*, 2004). The Ali-B model estimated total milk yield to within 2% of the actual milk yield. However, the prediction of total yield using the Ali and Schaeffer model deviated from actual total milk yield by almost 4%. The poor performance of Ali and Schaeffer's model in predicting the total milk yield is attributed to severe multicollinearity.

The Ali-B model has been found to be the most robust single equation model in this study. It is a relatively simple model, compared to the SLAC method that is currently being used, and therefore has many practical uses. However, while the parameters of Wood's model are biologically interpretable, those of the Ali-B model are not, as it is a polynomial equation, but this does not negate its proficiency. It is also concluded that the Ali-B model is the best at predicting weekly milk yield and it can therefore be easily used to create the seasonal production table for use by bio-economists.

The trend of the seasonal effects is very much as expected with the deviation in the late spring and early summer months being positive while in the winter months it is negative. This is similar to the findings of Killen and Keane (1978) but these effects are more extreme than those reported by Wood (1969), which would be expected bearing in mind that feeding regimes vary less throughout the year in the UK than in Ireland.

Adding the seasonal affects to the lactation curve derived using the Ali-B model, allows estimation of the percentage of the total lactation yield on each month throughout a lactation from a cow calving at any time throughout the year. As there was no significant difference between the parameters of the Ali-B model for each lactation number, a single equation model is adequate. However, lactation number has a significant effect on total milk yield and it was found, using analysis of variance, that as lactation number increased from one to two the total milk yield increased by approximately 16.2%, while between lactation numbers two and three (and greater) yield increased by approximately 6.6%; these findings are similar to those of Crosse, Van Heijst and O'Donovan (1988) and Buckley (1998). Therefore, to calculate milk production for a particular animal for each production month, the Ali-B model is used to calculate the weekly yield, and the production month defines what percentage of the cow's total milk yield is supplied in each month of the year (from seasonality production table).

Conclusions

It has been possible to arrive at a single, well-fitting and robust model to represent the shape of the lactation curve in Irish dairy herds. A number of previously derived models were examined and a modified version of the Ali and Schaeffer model was found to be the most satisfactory on the basis of its Mean Square Prediction Error value and its ability to satisfy the underlying assumptions of the regression procedure. This is a relatively simple model when compared to the SLAC method currently in use. The Ali-B model has four parameters and can be fitted to any data set using non-linear regression. When using this model to predict the milk yield for a specific cow, adjustments are made to account for variation attributable to seasonal effects on production. These effects may vary from region to region accounting for variation in factors such as climate, soil quality and environment. This model is also suitable for use by bio-economists who are constantly updating and re-creating the parameters for different scenarios. In conclusion, the Ali-B model is the best fitting model to Irish data and it can be easily updated for different regional effects.

Acknowledgements

The authors would like to acknowledge Mr. David Cliffe, Teagasc, Moorepark for his technical assistance and Mr. Gary Keogh, DCU, who gave advice on the analysis of residuals.

References

Ali, T.E. and Schaeffer, L.R. 1987. Accounting for covariance among test day milk yields in dairy cows. *Canadian Journal of Animal Science* **67:** 637–644

Belsley, D., Kuh, E. and Welsch E. 1980. "Regression Diagnostics: Identifying Influential Data and Sources of Collinearity". John Wiley and Sons,

Brody, S., Ragsdale, A.C. and Turner, C.W. 1923. The rate of decline of milk secretion with the advance of the period of lactation. *Journal of General Physiology* **5:** 441–444

Brody, S., Turner, C.W. and Ragsdale, A.C. 1924. The relation between the initial rise and the subsequent decline of milk secretion following parturition. *Journal of General Physiology* **6:** 541–545

Brotherstone, S., White, I.M.S. and Meyer, K. 2000. Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Animal Science* **70:** 407–415

Buckley, F. 1998. Effect of Genotype of Cow and system of Feeding on Milk Production, Feed Intake and Nutrient Utilisation of Spring-Calving Holstein-Friesian Dairy Cows in Ireland. Ph.D. Thesis, University College Dublin.

Carvalheira, J., Pollak, E., Quaas, R. and Blake, R. 2001. Fitting multiple lactations test-day records with an autoregressive repeatability animal model 12. *In proceedings of the 52nd* EAAP meeting, page 6.

Cobby, J.M. and Le Du, Y.L.P. 1978. On fitting curves to lactation data. *Animal Production* **26:** 127–133

Crosse, S., 1986. 'The Development and implementation of a computerized management system for Irish Dairy farmers'. PhD Thesis. National University of Ireland, Cork

Crosse, S., Van Heijst, G. and O'Donovan, S. 1988. The relative accuracy of different sampling procedures for estimating total lactation yield of milk and milk constituents. *Irish Journal of Agricultural Research* **27:** 31–38

Cunningham, E.P., 1972. Components of variation on dairy cow production. *Irish Journal of Agricultural Research* **11:** 1–10

Dave, B.K. 1971 First lactation curve of the Indian water buffalo. *JNKVV Research Journal* **5:** 93–98

Dechow, C., Rogers, G., Klei, L. and Lawlor, T. 2004. Heritability and correlations for body condition score and dairy form within and across lactation and age. *Journal of Dairy Science,* **87:** 717–728.

Grossman, M. and Koops, W. 1988. Multiphasic analysis of lactation curves in dairy cattle. *Journal of Dairy Science*, **71:** 1598–1608.

Guo, Z. and Swalve, H.H. 1995. Modelling of lactation curve as a sub-model in the evaluation of test day records. Paper presented at the INTERBULL open meeting, 7–8 September, 1995, Prague, Czech Republic.

Jamrozik, J., Ginola, D. and Schaeffer, L. 2001. Bayesian estimation of genetic parameters for test day records in dairy cattle using linear hierarchical models. *Livestock Production Science,* **71:** 223–240.

Jones, T. 1997 Empirical Bayes prediction of 305-day milk production. *Journal of Dairy Science,* **80:** 1060–1075.

Kellogg, D.W., Urquhart, N.S. and Ortega, A.J. 1977. Estimating Holstein lactation curves with a gamma curve. *Journal of Dairy Science* **60:** 1308–1315

Keown, J.F. and Van Vleck, L.D. 1972. Extending lactation records in progress to 305-day equivalent. *Journal of Dairy Science* **56:** 1070–1079

Killen, L. and Keane, M. 1978. The shape of lactation curves in Irish dairy herds. *Irish Journal of Agricultural Research* **17:** 267–282.

Koonawootrittriron, S. Elzo, M. A., Tumwasorn, S. and Sintala, W. 2001. Lactation curves and prediction of daily and accumulated milk yields in a multibreed dairy herd in Thailand using all daily records. *Thai Journal of Agricultural Science* **34:** 123–139.

Kvanli AH. Guynes C.S. and Pavur R.J. 1986. Introduction to Business Statistics. Fourth Edition. West Publishing Company.

Lidauer, M., Mantysaari, E., Stradem, I. And Poso, I. 2000. Multiple-trait random regression test-day model for all lactations. *In the proceedings of INTERBULL meeting,* pages 81–86.

Madalena, F.E., Martinez, M.L. and Freitas, A.F. 1979. Lactation curves of Holstein-Friesian and Holstein-Friesian x Gir cows. *Animal Production* **29:** 101–107

Maddala, G.S. 1992. 'Introduction to Econometrics'. 2nd Edition. Macmillan Publishing Company, New York.

Mendoza, R.U. 1999. Serial Autocorrelation: Why it's there, what it does, and how to get rid of it. Fordham University. New York.

Molina, J.R. and Boschini, C. 1979. Adjustment of dairy curve of holstein herd with a linear modal model. *Agronomia Costarricense* **3:** 167–174

Olori, V.E., Brotherstone, S., Hill, W.G. and McGuirk, B.J. 1999. Fit of standard models of the lactation curve to weekly records of milk production of cows in a single herd. *Livestock Production Science* **58:** 55–63

Olori, V.E. and Galesloot, P.J.B. 1999. Projection of partial lactation records and calculation of 305–day yields in the Republic of Ireland. *Proceedings of the 1999 Interbull Meeting* Zurich, Switzerland, August 26–27, Pages 149–154.

Pérochon, L., Coulon, J.B. and Lescourret, F. 1996. Modelling lactation curves of dairy cows with emphasis on individual variability. *Animal Science* **63:** 189–200

SAS(r) Online Doc, Version 8, 1999. SAS Inst., Inc. Cary NC.

Schaeffer, L. and Burnside, E. 1976. Estimating the shape of the lactation curve. *Canadian Journal of Animal Science* **56:** 157–170

Sen, A. and Srivasta, M. 1990. 'Regression analysis-Theory, Methods and Applications.' Springer-Verlag. New York.

Shanks, R.D., Berger, P.J., Freeman, A.E. and Dickinson, F.N. 1981. Genetic aspects of lactation curves. *Journal of Dairy Science* **64:** 1852–1860

Shalloo, L., Dillon, P., Rath, M. and Wallace, M. 2004. Dairy farm simulation model description and validation of the Moorepark dairy system. *Journal of Dairy Science. [In Press]*

Sherchand, L., McNew, R., Kellogg, D. and Johnson, Z. 1995. Selection of a mathematical model to generate lactation curves using daily milk yields of Holstein cows. *Journal of Dairy Science* **78:** 2507–2513

Sikka, L.C. 1950. A study of lactation as affected by heredity and environment. *Journal of Dairy Research* **17:** 231–252

Singh, R.P. and Gopal, R. 1982. Lactation curve analysis of buffaloes maintained under village conditions. *Indian Journal of Animal Sciences* **52:** 1157–1160

Vasconcelos, J., Martins, A., Petin-Batista, M., Silvestre, A., Ferreira, A., Colaao, J. and Carvalheira, J. 2002. Estimation of lactation yields from test day yields of milk, fat and protein for Portuguese Holstein cattle using autoregressive test day model. *In proceedings of the 7th* World Congress on Genetics Applied to Livestock Production, Montepellier, France. pages 279–282,

Wilmink, J.N.M. 1987 Adjustment of test day milk, fat and protein yield for age, season and stage of lactation. *Livestock Production Science* **16:** 335–348

Wood, P.D.P. 1967. Algebraic model of the lactation curve in cattle. *Nature, London* **216:** 164–165

Wood, P.D.P. 1969. Factors affecting the shape of the lactation curve in cattle. *Animal Production.* **11**: 307

Yadav, M.P., Katpatal, B.G. and Kaushik, S.N. 1977. Components of inverse polynomial function of lactation curve, and factors affecting them in Harina and its Friesian crosses. *Indian Journal of Animal Sciences* **47:** 777–781