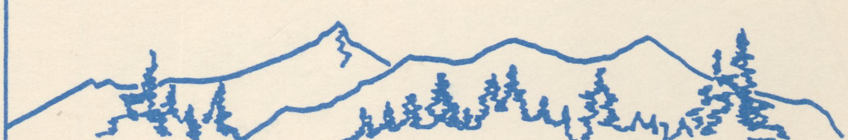
Papers of the 1991 Annual Meeting

Western Agricultural Economics Association



Portland, Oregon
July 7 – 10, 1991

Title: Lactation Curve Estimation for use in Economic Optimization Models in the Dairy Industry

Authors: Timothy J.Richards
Univ. of Lethbridge
Brian Freeze
Agriculture Canada

Abstract: A three stage least squares lactation curve model is estimated for milk production, fat content, protein content and body weight change in lactating Holstein cattle. The study found that the simultaneous procedure is preferred both theoretically and empirically for constructing lactation curves relevant for use in dairy economic optimization models.

I). Introduction

Profit maximizing mathematical programming models of dairy production need to recognize the complex nature of milk production. Individual cows have differing inherent abilities to convert feed to milk depending upon their genetic backgrounds, their response to environmental conditions, age, breed, stage of lactation and other intangible factors. Independent of any management decision, milk and milk component production changes over time due to the biology of a lactating animal. In order develope a model which maximizes the profit from marketing milk over time, the manager must develope an accurate model of how his decisions regarding breeding, calving, feeding and culling affect the structure of the lactation curve, milk component curves and weight change curve. These relationships are required to predict the nutritional requirements for a specific animal based upon her individual performance characteristics.

The objective of this study is to use monthly data on completed lactations to estimate lactation curve parameters to be used within a profit maximizing mathematical programming model. In comparison to previous lactation curve work, the current study employs a different algebraic model of the lactation curve, a simultaneous rather than single equation estimation technique, monthly rather than daily or weekly observations, and a more explicit treatment of the genetic background of individual animals. Besides controlling for genetic effects and dietary effects on the lactation curve, the current model attempts to isolate the seasonal effect of calving date and current production month as well as the age of the cow. By allowing for the simultaneous explanation of the various measures of cow 'performance', the model will allow for a more realistic description of how diets are formulated in optimization models. Diet, production and weight changes are determined simultaneously in reality and not independent of one another.

Previously, dairy economic optimization models have incorporated lactation curves in a variety of ways. Reyes et al. (JDS, 1981) constructed optimal diets for six or seven stages for 36 yield/age/calving season combinations for either a 13 month or 15 month calving interval. Their lactation curves were taken from results in the existing literature for each calving season so do not refer to the population that they are constructing the optimization program for. Klein et al. (1986) used five lactation segments each eight weeks long. However, production and feed intake were determined experimentally for each stage and production potential level, so their results were not parametrically summarized in their paper. In a model of Western Australian Dairy farming, Olney and Falconer (1985) utilize a declining step function to approximate a lactation curve. They assume that each stage represents a certain percentage of the potential production in the previous stage and, therefore, do not rely upon empirically or scientifically determined results. Other studies ignore the fact that the profitability of an individual cow changes as her ability to convert feed to milk changes over the lactation (Kleyn and Gous, 1988; O'Connor, et al., 1989).

Most lactation curve studies employ a variation of, or compare their model results to, the incomplete gamma (IG) function originally estimated by Wood (1967). Algebraically, the deterministic incomplete gamma lactation curve function can be written as:

$$(1)\quad y(t) = a \cdot t^{b} \cdot \exp(-ct)$$

where y is the average amount of milk per day, t is the number of days after calving, a is a scale parameter, and b and c are both shape parameters for the lactation curve.

This function has become widely used because of several desirable characteristics. First, a simple logarithmic transformation of the above curve with a multiplicative error term can be estimated by multiple linear regression:

$$(2)\ \log(y(t)) = \log(a) + b\cdot\log(t) - c\cdot t + \log(\epsilon)$$

where ϵ is assumed to be normally distributed with zero mean and finite variance. Secondly, several important values summarizing the shape of the lactation curve can be calculated from the above parameter estimates. The number of days to peak yield (t*) can be calculated by $t^*=b/c$. Integrating the function over the lactation will yield the total amount of milk produced:

$$(3)\ y = (a/c^{t+1})\Gamma(b+1)$$

In addition, a measure of persistency of production over the lactation can be expressed as $S = c^{-(b+1)}$. Although this measure is not in units convenient for interpretation, it can be usefully employed in comparing persistency between estimated lactation curves.

However, several researchers have noticed problems in estimating this function in practice. Most notably, Dhanoa and Le Du (1982, DLD); Cobby and Le Du (1978, CLD); Goodall and Sprevak (1984, GSa) and DeLuyker et al. (1990, DSWAW) found that the residuals from fitting the IG curve were serially correlated in their data. This finding, in addition to the biased parameter estimates claimed by Congleton and Everett (1980, CE) and Rowlands et al. (1983, RLR) suggests that alternative model specifications could potentially be used to improve the goodness of fit of the basic Wood model. Many of the above studies (DLD; GSa,b; DSWAW), use the serially correlated residuals as an indication of the ability of time series or distributed lag methods to improve the model. Others interpret these failures as a need to search for an inherently different model structure. CLD improved on the fit of the Wood lactation curve model through the use of a model with a linear decline in production with time. Recently, Grossman and Koops (GK) have succeeded in removing autocorrelation in the residual series through the use of diphasic and triphasic functions. These functions use accumulated production across phases to explain the shape of the lactation curve. Unsatisfied with the inconsistencies in comparing the success of the Wood model with the CLD model lead Elston et al. (1989, EGN) to develope a non-parametric methodology. These studies suggest that individual lactation curves are highly dependent upon the nature of the data available and the particular herd being analyzed.

In the current study, it was found that a polynomial log functional form performed the best in explaining the variation in milk production across the lactation. This function maintains much of the flexibility of the incomplete gamma function while limiting the restrictions placed upon the empirically determined shape in any given data set. Since our data set consists only of monthly observations on milk volumes, we sought to avoid a functional form that is sensitive to few observations prior to peak lactation time. Congleton and Everett (1980) find the IG curve to frequently yield degenerate curve estimates (ie. negative days to peak yield) when pre peak observations are few. Including a multiplicative error term and the other exogenous, pre-determined and endogenous variables in exponential form yields the equation:

$$(4)\ M(t) = a\cdot t^{b}\cdot \exp[c(\log t)^2]\cdot \exp[dX]\cdot \epsilon_t$$

where M is daily milk production, t is days post partum; X is the matrix of exogenous, pre-determined and joint endogenous variables; a,b and c are parameters; d is a vector of parameters and ϵ is a zero mean, finite variance, serially uncorrelated error term. This function yields the following log transformation which allows for simple linear multiple regression:

$$(5)\ \log(M(t)) = \log(a) + b\log(t) + c(\log(t))^2 + dX_m + \log\epsilon$$

From this equation, the time to maximum yield can be calculated from the estimated parameters as $t^* = \exp(-b/2c)$. The total yield over the lactation can be calculated and predicted through numerical methods of integration. Because milk production ability is also believed to peak in age, the age variable was included in equation (5) in a quadratic manner.

Wood (Wb, 1976) and Morant and Gnanasakthy (MG, 1989) apply conventional models of lactation curves to study the change of component concentration and yields over time. However, much of the change in total yield can be explained by the change in the volume of carrier over the lactation, so this procedure does not add significantly to wheat we know about the change in the ability to produce fat and protein over time. In the data used by Wood, however, the IG model appears to fit concentration levels very well. Zanartu

(1983) presents experimental results on how component concentration changes over time. Based upon these empirical observations, alternative functional forms were chosen for the fat and protein concentration equations. Specifically, they found that fat percentage began at a relatively high level and declines geometrically as the lactation progresses, contrary to the results of Wb. Therefore, fat content was modeled as an inverse function of the days post-partum as represented by the following equation:

$$(6)\ \text{Fat\%}(t) = f + g/t + hX_f + \mu$$

where X_f is the vector of right hand side variables particular to the explanation of fat content and μ is a vector of zero mean errors.

Zanartu et al. also found that protein concentration changed over the lactation as well. Beginning at a moderate level, the protein percentage increases at an increasing rate with days after calving. To model this type of relationship, the current study treats protein as a function of lactation days squared, in addition to the conditioning variables specific to the protein equation:

$$(7)\ \text{Protein \% }(t) = j + k\cdot t + n\cdot t^2 + q\cdot X_p + \tau$$

With respect to weight change, the algebraic structure of the model was simply linear in the untransformed variables. From experimental results found by Zanartu et al., there does not appear to be any discernable shape to the change in weight over the lactation beyond a simple linear trend.

Variables under management control, such as breeding and feeding, have been shown through experimentation to affect the levels of production, concentration and weight change of lactating cows. Since diet formulation and genetic background both have a significant effect on the animal performance criterion, the lactation curve model must control for these influences.

Independent of the stage of the lactation, Sutton (JDS, 1989) concluded, upon review of the nutrition literature, that nutritional means can be used to manipulate the concentration of the components found in milk. Many experiments have shown that diets involving extremely restricted or unbalanced formulations can affect milk composition. Particularly, diets low in forage (ie. below 50%) can cause milk fat depression of a magnitude of 1.5% or better. The most important dietary determinant of milk fat composition was shown by Sutton (JDS, 1989) to be the concentrate dry matter intake. In addition, with constant energy, milk fat depression was associated with a rise in milk yield. Zanartu, et al. (1983) found that a diet balanced according to NRC requirements, termed restricted concentrate, led to lower milk protein on average and a reduced tendency for milk protein to increase as the lactation progressed. Following the NRC diet lead to both lower bodyweight and fat content throughout the lactation.

Besides the effect of diet on milk composition, an individual producer can modify the composition and yield of milk through genetic selection. Data presented in Gibson (JDS, 1989) supports the contention that breeding for performance will be effective, particularly in selecting for fat:protein ratio. Although the results cited for heritability and genetic correlation of the relevant performance factors are lower than one might expect, the value are only averages and not derived from complete models designed to isolate genetic effects. Controlling for all other relevant variables, the effect of genetic selection suggested by Gibson would be more conclusive. Shanks et al. (1981) found a significant degree of genetic correlation between peak yields using an IG lactation curve model. Controlling for breed and sire effects, Grossman et al. (GKN, 1986) were also able to differentiate between lactation curves on a genetic basis. Using data on Canadian Holsteins, Batra (1986) also found significant differences between lactation curves for different dam and sire lines, using both the IG curve and the inverse polynomial function.

In order to control for the effects of differing genetic potentials between animals, the current study utilizes data on the 'proof' of each cow's sire. Although this includes only part of the genetic background of the cow, it should capture the dominant effect. To allow for differences in diet, lactation curves were constructed for different feeding groups in the herd.

III). Statistical Model

For this study, data was obtained on 137 completed lactations using the University of Saskatchewan Holstein dairy herd. The sample consists of first, second, third and fourth lactation results. Monthly observations on daily average milk production, fat concentration, protein concentration, and weight change were obtained for each lactation. For each of these observations, data was also obtained on the calving date, the age of the cow, the cow's sire proof, bodyweight and the animal's feeding group for the month. The sire data was obtained from the Canadian Holstein Association.

The model designed to test the effects of the exogenous variables on performance factors and the interaction effects between factors was as follows:

$$(1) \qquad Y_i\Gamma_i + X_i\theta_i + E_i = 0$$

where Y_i is a T x M matrix of T observations on the M jointly endogenous or animal performance variables for the ith feeding group;
X_i is a T x K matrix of T observations on the K exogenous or explanatory variables for the ith feeding group;
Γ_i is a conformable coefficient matrix on the M endog variables;
θ_i is a conformable coefficient matrix on the K exog variables;
E_i is a matrix of random error terms with the characteristic:

$$E = (e_1\ e_2\ e_3\ .\ .\ .\ e_m) \text{ such that } E[ee'] = \Sigma \quad I_T$$

and Σ is the contemporaneous covariance matrix which is assumed to be symmetric, positive semidefinite and, therefore, non-diagonal.

Each system of four equations was estimated separately for the ten feeding groups considered. The feeding groups were based upon differing nutritional requirements based upon health, pregnancy, performance or other reasons. In order to consistently estimate the parameters in the presence of simultaneity, an instrumental variables procedure was required. Within this class of estimator, a full information estimator was required to take advantage of the non-diagonal cross equation covariance matrix. Therefore, three stage least squares was used for each system to produce asymptotically efficient and unbiased parameter estimates.

In the current model, the jointly endogenous animal performance factors consist of average daily milk production per month, average milk fat concentration per month, average protein concentration per month and average daily weight change per month.

In the design of the structural equation system, allowances were made for the joint endogeneity of the performance variables. Since milk production, component concentration and weight change are all affected simultaneously by the diet and other management variables, the interaction effects between them must be accounted for in a statistical model of performance. For example, Sutton (1989) notes that "... depression of milk fat concentration on low roughage diets is usually accompanied by a reduction in milk fat yield and a consequent increase in energy retention or live weight gain, emphasizing the importance of considering the effects of such diet changes on the overall energy balance of the animal rather than on milk production alone". Empirical results such as this suggest that the level of each endogenous variable changes only as components of the larger system change and that changes in one variable cause repercussions throughout all of the performance-requirements relationships. In order to capture seasonal production effects, only the months of March to August were included as qualitative exogenous variables. These months reflect both the natural tendency of cows to produce more in the spring months as well as the stress reduction from the milder temperatures. SWMT show climactic factors to be important variables in determining the lactation structure of milk cows, so the inclusion of the non-winter months acts as a proxy variable for the climate effect, given our lack of detailed environmental data. Since pasturing plays a minor role in herd management, the seasonality of grass feed should not cause this variable to be as important as in other studies (ie. RLR). Including current month may, however, control for the effect of the new quota year (August) and the corresponding incentives to producers to either increase production or cut back if currently over quota.

Table one presents the estimated coefficients for feeding group number one for all system equations. Since space limitations prevent the printing of all equations' results, feeding group one is presented as it represent the largest sample size within the Saskatchewan data set.

<u>Table One: Group One Lactation Curve Parameters</u>

Equation

variable	milk	fat	protein	weight change	
CONSTANTLL	0.057		2.072*	-5.037* -	-5.235*
ACDAY	1.490*		-----	----	-----
LLACDAY2	0.189*	--	-----	-----	-----
LACDAY	-----		-----	-----	0.004*
LACDAY2	------		-----	0.001*	-----
INVLACDACA	------		3.234*	-----	-----
VMON1	-0.036		0.056	-0.020	0.282
CAVMON2	0.030		0.080	-0.055	0.090
CAVMON3	0.009	-	0.017	0.013	0.157
CAVMON4	0.007		0.031	-0.034	0.323*
CAVMON5	0.019		0.023	-0.009	0.272
CAVMON6	-0.064		-0.015	0.030	0.313
CAVMON7	-0.128*		-0.004	0.066	0.288
CAVMON8	0.010		-0.011	0.016	0.243
CAVMON9	-0.057		0.147*	-0.114*	0.316
CAVMON10	-0.005		0.021	-0.003	0.147
CAVMON11	0.001		-0.004	0.050	-0.045
MONTH3	-0.053		-0.018	-0.007	0.145
MONTH4	-0.123*		-0.034	-0.047	0.617*
MONTH5	-0.019		-0.033	0.021	-0.043
MONTH6	0.018		-0.057*	0.021	-0.057
MONTH7	0.004		-0.075*	0.026	-0.046
MONTH8	-0.002		-0.034	-0.021	0.084
AGE	0.005*		0.008*	-0.015*	-0.032
AGESQ	-0.014*		-----	-----	-----
MILKPF	0.180*		-----	-----	-----
FATPF	-----		0.004*	-----	-----
PROTPF	-----		-----	0.004	-----
TYPEPF	-----		-----	-----	0.004
MILK	-----		-0.835*	1.994*	1.672*
FAT	-----		-----	0.800*	-0.758
PROTEIN	-----		0.687*	-----	-----
WGTCHANGWE	0.138*		-0.013	0.138*	-----
IGHT	0.003		-----	-----	-----
R2	0.507		0.638	0.911	0.075

* indicates significant t statistic at 5% level

From the polynomial log milk equation, the estimated parameters suggest that milk production peaks at 51.8 days into the lactation. Compared to the results obtained by Sharma et al. (SWMT, 1986), these results accord very closely to the results achieved in practice. Milk production also attains a maximum in age at roughly 6.5 years. Again, this result accords well with actual results. For this feeding group, the equation explained 50.7% of the variation in the log of milk production. Compared to the transformed Wood equation's performance in our data, the polynomial log explained 7% more variation in production. Significant seasonal effects include only calving in July and production in April. Lactation curves for animals with a higher genetic potential to produce milk are slightly higher, as we would expect. Empirically, weight change over the lactation is positively correlated with milk production.

Modeling butterfat as inversely related to lactation stage is supported by the current data set. With t the current lactation day, fat content is given by FAT % = 2.072 + 3.234/t. As in the milk equation, these figures are quite consistent with practice. In addition, older cattle producing a higher level of protein and less milk with more potential to produce fat are associated with higher fat concentrations. This equation explains 64% of the variation in fat content.

Although the R^2 value is the highest for the protein structural equation, the parameters are less believeable than for the previous equations. According to this equation, protein content starts from a negative level and increases with 0.001 in days squared. Protein declines with age, but increases strongly in milk production, fat production and weight change. Changes in bodyweight remain 92.5% unexplained by this model, but still increase significantly in milk production, when all other management and environmental factors are held constant. Just as with the other performance variables, the lactation phase explains a large proportion of the physiological condition of the animal.

IV). Discussion

This study found significant empirical differences between lactation curves of cows that differ in age, genetic potential, and production levels of other milk components and body tissue. Certain seasonal effects were also isolated. The findings support modeling lactation curves as simultaneously determined with diet and other performance objectives.

Estimated parameters from this model indicate different economic tradeoffs from those currently employed in optimization models. Specifically, the positive correlation of milk with weight gain suggests that feeding for a weight loss-milk production tradeoff may not be as economically sound as previously thought. With a protein pricing system, a clear incentive exists to increase the volume of milk production, due to the strength of the interaction of these two variables. With this relationship in mind, breeding for protein could be used to mitigate the fat depression associated with high milk production, given the positive bilateral relationship between fat and protein content. As protein levels increase with the weight change, a protein pricing scenario would further strengthen the argument for maintaining body condition across the lactation. Considering performance factors as simultaneously determined greatly enhances the economic content of the coefficients describing productivity changes over the lactation.

Based upon the favorable results of SWMT, GKN, Batra (1986) and others, the results of this study would be improved with the inclusion of a measure of pregnancy status of the lactating cow. Increased estimation efficiency could be achieved both with more frequent milk recording data as well as a greater number of completed lactations.

V). References

Batra, T.R. 1986. "Comparison of two mathematical models in fitting lactation curves for pureline and crossline dairy cows". Canadian Journal of Animal Science. 66 p.405.

Cobby, J.M. and Y.L.P. Le Du. 1978. "On fitting curves to lactation data". Animal Production. 26. p.127.

Congleton, W.R. and Everett, R.W. 1980. "Error and Bias in using the incomplete gamma function to describe lactation curves". JDS. 63. p.101

DeLuyker, H.A.; R.H. Shumway; W.E. Wecker; A.S. Azari; and L.D. Weaver. 1990. "Modeling daily milk yield in holstein cows using time series analysis". JDS. 73. p.539

Dhanoa, M.S. 1981. "A note on an alternative form of the lactation model of Wood". Anim. Prod. 32. p.349

----------- and Y.L.P. Le Du. 1982. "A partial adjustment model to describe the lactation curve of a dairy cow". Anim. Prod. 34. p.243.

Elston, D.A.; C.A. Glasbey and D.R. Neilson. 1989. "Non-parametric lactation curves". Anim. Prod. 48. p.331.

Goodall, E.A. and D. Sprevak. 1984. "A Note on a stochastic model to describe the milk yield of a dairy cow." Anim. Prod. 38. p.133

---------- . 1985. "A Bayesian estimation of the lactation curve of a dairy cow". Anim. Prod. 40. p.189.

Grossman, M. and W.J. Koops. 1988. "Multiphasic analysis of lactation curves in dairy cattle". JDS. 71. p.1598.

---------- ; A.L. Kuck and H.W. Norton. 1986. "Lactation curves of purebred and crossbred dairy cattle." JDS. 69. p.195

Klein, K.K.; R. Hironaka; C.H. Heller and B.S.Freeze. 1986. "Profit Maximizing linear program model for dairy cattle". JDS. 69. p.1070

Morant, S.V. and A. Gnanasakthy. 1989. "A new approach to the mathematical formulation of lactation curves". Anim. Prod. 49. p. 151.

O'Connor, J.D.; C.J. Sniffen; D.G. Fox; and R.A. Milligan. 1989. "Least Cost dairy cattle ration formulation model based on the degradable protein system". JDS. 72. p.2733.

Olney, G.R. and D.A. Falconer. 1985. "Mathematical Programming model of Western Australian dairy farms". Western Australian Dept. of Agriculture. South Perth, W.A.

Reyes, Alberto A.; Robert W. Blake; C. Richard Shumway; and James T. Long. 1981. "Multistage optimization model for dairy production". JDS. 64. p. 2003.

Rowlands, G.J.; S. Lucey and M. Russell. 1982. "A comparison of differerent models of the lactation curve in dairy cattle". Anim. Prod. 35. p.135.

Sharma, A.K.; C.J. Wilcox; F.G. Martin and W.W. Thatcher. 1990. "Effects of stage of lactation and pregnancy and their interactions on milk yield and constituents". JDS. 73. p.1586.

Wood, P.D.P. 1967. "Algebraic model of the lactation curve in cattle". Nature. 216. p.164.

---------. 1976. "Algebraic models of the lactation curves for milk, fat and protein production with estimates of seasonal variation". Anim. Prod. 22. p.35