Modelling Liveweight Change over a Lactation in Irish Dairy Cows

Noreen Quinn[1,2], Lynn Killen[1] and Frank Buckley[2]
[1] School of Computing, Dublin City University, Dublin 9, Ireland,
[2] Dairy Production Research Centre, Teagasc, Moorepark, Fermoy, Co.Cork, Ireland

Abstract

The aim of this study is to derive an equation that has the ability to model liveweight change of Irish dairy cows over a lactation period. The dataset consists of 11,055 liveweight recordings taken at monthly intervals and 1,446 at daily intervals for spring calving cows from 83 herds together with 6,893 liveweight recordings from 2 autumn herds. An initial examination focussed on time series techniques, as the data is of a time series nature. Splines were also examined to determine the dimensions of a model required to represent the data. As an incomplete gamma function, which was previously used to model milk yield, has been used in other studies to model liveweight, various milk yield models were investigated. Finally, liveweight changes between two calvings were modelled as a function of age, lactation and pregnancy. As multicollinearity was evident in this function, the variance inflation factor was examined and principal component analysis was carried out on the variables responsible for multicollinearity. The new liveweight model has a much better fit than previous models, weak multicollinearity and the residuals are homoskedastic, independent and normally distributed. This liveweight model therefore provides an acceptable level of accuracy in representing the shape of the liveweight curve for Irish dairy cows, and can be easily modified for different environmental scenarios.

Keywords: Liveweight change; Analysis of Residuals; Pregnancy; Principal Component Analysis; Splines; Seasonal effects.

Introduction

Accurate estimates of liveweight of individual animals can be beneficial when making management and nutritional decisions both at herd level and for individual cows (Forbes, 1983; Walter *et al.,* 1984). The Moorepark Dairy Systems Model (MDSM) (Shalloo et al., 2004) is a farm simulation model that requires an accurate description of the liveweight profiles of Irish cows. The loss and gain of liveweight has a net cost in energy within the production system, and it is necessary to include the change in liveweight in an economic model to accurately reflect the production system. A realistic model to estimate and predict liveweight change of an animal throughout the year is therefore worthy of investigation. Liveweight has been modelled using three approaches – modelling liveweight from birth to maturity (Brown *et al.*, 1976; Bakker and Koops, 1978; Taylor, 1980; Moore, 1985; Perotto *et al.,* 1992; Berry *et al.,* 2004), modelling liveweight using body measurements (Gravir, 1967; Heinrichs *et al.*, 1992; Wicks, 2001; Madalena *et al.*, 2003) and modelling liveweight over a lactation period (Wood *et al.*, 1980; Korver *et al.*, 1985; Berglund and Danell 1987; Lopez-Villalobos *et al.*, 2001). The focus of interest in this study is the evolution of the liveweight of a dairy cow throughout a lactation.

Many researchers have contributed to the continuous progression of work in modelling patterns in liveweight change of individual animals; Wood *et al.* (1980) examined the liveweight changes of several breeds of British dairy cows. They used an incomplete gamma function, which had previously been used to model milk yield; however, their analysis was restricted to the first 20 weeks after calving. Korver *et al.* (1985) constructed a function, from the incomplete gamma function, incorporating liveweight level (scale), a pregnancy parameter, the maximum decrease of liveweight during the lactation and the time during lactation at which minimum liveweight occurred. Berglund and Danell (1987) and Lopez-Villalobos *et al.* (2001) also used the model of Wood *et al.* (1980) to predict liveweight change.

The use of time series techniques was examined initially, as the data involved in this study are inherently of a time series nature. The dimensions of the model required to fit the liveweight data was approximated using splines. As cubic splines are the most widely used splines, they were invoked in this study. The principal model, to date, to describe liveweight over a lactation is the model of Wood *et al.* (1980). As this model form was previously used to describe milk yield, other models that were used by Quinn *et al.* (2004) to describe milk yield were also investigated. While the suitability of the models was primarily judged on the basis of "goodness of fit", a residual analysis was carried out to test the validity of the assumptions of regression analysis, namely autocorrelation, homoskedasticity, multicollinearity and normality of distribution of error terms. The effect of environmental and seasonal factors, free of stage of lactation, was also examined.

Materials and Methods

Data

The data used in this study comprised of 11,055 liveweight recordings taken at monthly intervals and 1,446 taken at daily intervals for spring-calving cows from 83 herds, and 6,893 liveweight recordings from two autumn-calving herds. Spring-calving cows were defined as cows that calved between January and June, while those calving in the remainder of the year were defined as autumn-calving. The liveweight recordings were collected, by the DairyMis system in Teagasc (Crosse, 1986) over the period 1995 to 2001. All herds in this study were representative of Irish pasture-based seasonal calving herds. The data included year of production, parity (lactation number), calving month, lactation week and liveweight. For the purposes of this study, lactation number was categorized as lactation 1, lactation 2 and lactation 3 or greater (Cunningham, 1972). Records that contained fewer than five recordings, records with no recording while the animal was pregnant and lactations of less than 25 weeks duration were removed. Liveweight was recorded by an automatic weighing system (DairyMaster); this system consisted of a scale with load cells and thus there is no visual recording of the weight. In all cases, recordings were taken after milking so as to minimize variations due to changing weights of gut fill.

Models and Statistical Analysis

The dimensions of the liveweight data were investigated, in an initial examination, using cubic splines. A cubic spline is a third-order curve applied to a set of *m* control points. If there are one or more splines, the abscissa values of the join points are called knots. The general form of a third-order or cubic polynomial is given by:

$$f(x) = ax^3 + bx^2 + cx + d$$

where *a, b, c* and *d* are constant coefficients. A condition of a cubic spline is that its derivative and its second derivative are continuous at the knots and that its second derivative is commonly set to zero at the endpoints since it provides a boundary condition that completes the system of m-2 equations. By subtracting the number of continuous derivatives from the total number of degrees in the spline, the dimensions of the dataset are calculated.

Once the dimensions of the data were determined, the problem of deriving an equation to represent the data could be explored. The models under consideration, namely Wood et al. (1980), Wilmink (1987) and Guo and Swalve (1995), were tested on the basis of their goodness-of-fit and their ability to adhere to the assumptions of regression analysis. The Mean Square Prediction Error (MSPE) value was used as a measure of goodness-of-fit (Kvanli, Guymes and Pavur, 1986). The Durbin-Watson statistic, d, was calculated for each model to test for the existence of autocorrelation between the residuals; the decision rules for autocorrelation used in this study are those outlined by Mendoza (1999). Initially, first order autocorrelation was examined and if this proved to be inconclusive, higher order autocorrelations then were tested. A condition index was calculated to test for the presence of multicollinearity, values ranging from 30 to 100 indicating that moderate to strong multicollinearity was present (Belsley, Kuh and Welsch, 1980). Multicollinearity gives rise to two problems: 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 should be removed but if this is insufficient, principal component analysis (PCA) can be utilised. White's test was used to test for heteroskedasticity for each individual lactation and a mean value was computed after accounting for calving month, lactation number and herd effect. Additional tests included the Kolmogorov-Smirnov statistic (D), a test for normality of the distribution of the residuals as well as tests for kurtosis (weight in the extremes) and skewness (lack of symmetry).

PCA is a technique for forming new variables that are linear components of the original variables. The maximum number of new variables that can be formed is equal to the number of original variables. The first principal component (or new variable) accounts for as much of the variability in the data as possible and each succeeding component accounts for as much of the remaining variability as possible. The normal convention is to standardise the data before carrying out principal component analysis so that each recording makes an equal contribution to the total variance. Finding the principal components for two variables involves an orthogonal rotation of the axes. The first principal component will be in the direction of greatest variance and this is found by minimising the sum of the squared perpendicular distances from the observations to the first component. Once the first component is positioned, the second component is fixed since it must be orthogonal to the first. The principal components are, as a result, uncorrelated among themselves.

Once a model was identified, the deviations found by comparing the each data point with the corresponding value as estimated by the model were cumulated for each calendar month, and the mean of the deviations was computed. These mean deviations estimated the effect of some environmental factors free of stage of lactation effect.

Results and Discussion

The preliminary examination dismissed the use of time series techniques because these techniques require that data points occur at equal time intervals, which is not the case with these data. The examination of splines showed that two cubic splines with one knot best fitted the data. Thus, the model required to represent the liveweight curve would be a four dimensional equation, as two cubic splines, (one before and one after the knot) has a total degree of six and it involves two derivatives, which reduces the dimensions of the equation to four.

Models which were examined by Quinn *et al.* (2004) to describe the evolution of milk yield, were fitted over a lactation to the liveweight data. The models of Quinn *et al.* and Ali and Schaeffer were eliminated as they are polynomial expressions and thus keep their concave shape. The analysis of residuals showed that there was strong multicollinearity with the model of Wood *et al.* (1980) and Guo and Swalve (1995) (Table 1); however the model of Guo and Swalve (1995) had the best MSPE value.

Table 1: *Comparison of Models cited in Literature*

Test	Wood	Wilmink	Guo & Swalve
MSPE	311.45	316.27	308.98
R^2	0.62	0.58	0.60
Autocorrelation	No	No	No
(Durbin-Watson)	1st Order	1st Order	1st Order
Heteroskedasticity	No	No	No
(White's Test) [p-value]	0.33	0.32	0.33
Normality	Normal	Normal	Normal
(Kolmogorov-Smirnov)[p-value]	0.18	0.17	0.17
Kurtosis	1.18	1.06	1.16
Skewness	-0.04	0.02	-0.03
Multicollinearity	Strong	Moderate	Strong
(Condition Index)	88.67	25.98	88.16

Other factors effecting liveweight, such as age and pregnancy, were also examined and it was concluded that liveweight changes of a dairy cow could be modelled as a function of age, lactation and pregnancy in the following way:

$$LW_n = f(age) + f(lactation) + f(pregnancy)$$

where LW_n= the liveweight in lactation week *n*. As dairy farmers in Ireland operate a strict calving pattern, the age at calving within lactation does not vary to any great degree and thus, a constant multiplied by lactation number was considered to be appropriate as the function of age. As all of the models described in this analysis were functions of lactation, it was decided that the model that provided the best fit to the data would be chosen namely that of Guo and Swalve (1995). The function described by Huggett and Widdas (1951) to represent the effect of pregnancy on liveweight was incorporated into our model. Thus the total function describing the combined effects of age, lactation and pregnancy on liveweight is as follows:-

$$LW_n = \underset{\substack{\downarrow\\ a(lactation\ number)}}{f(age)} + \underset{\substack{\downarrow\\ b + c\sqrt{n} + d}}{f(lactation)} + \underset{\substack{|\\ g(days\ pregnant - h)}}{f(pregnancy)}$$

where LW_n= the liveweight in lactation week *n* and *a, b, c, d, g* and *h* are parameters. As the lactation number is constant for each record, the function of age was combined with the constant term to give the following model:-

$$LW_n = a + c\sqrt{n} + d\ln(n) + g(days\ \ pregnant - h)^3$$

The effect of multicollinearity was evident in this function and therefore the variance inflation factor was examined to determine which variables were correlated. The terms $\sqrt{n}$ and $\ln(n)$ were highly correlated with variance inflation factor values of 25.88 and 22.58, respectively. Thus, PCA was carried out to replace these two terms that were correlated with two independent linear components:-

$$\sqrt{n} = \alpha_{11}PC1 + \alpha_{12}PC2$$
$$\ln(n) = \alpha_{21}PC1 + \alpha_{22}PC2$$

where *PC1* and *PC2* are principal component scores one and two, respectively and α_{ij} are the eigenvectors associated with the i^{th} variable and the j^{th} principal component. These two independent linear components describe all the variation in the two original variables leading to the following functional form:-

$$\begin{aligned} LW_n &= a + c[\alpha_{11}PC1 + \alpha_{12}PC2] + d[\alpha_{21}PC1 + \alpha_{22}PC2] + g(m-h)^3 \\ &= a + (c\alpha_{11} + d\alpha_{21})PC1 + (c\alpha_{21} + d\alpha_{22})PC2 + g(m-h)^3 \\ &= \beta_0 + \beta_1 PC1 + \beta_2 PC2 + \beta_3 (m-h)^3 \end{aligned}$$

where $\beta_0 = a$, $\beta_1 = (c\alpha_{11} + d\alpha_{21})$, $\beta_2 = (c\alpha_{21} + d\alpha_{22})$, $\beta_3 = e$, *PC1* and *PC2* = principal component scores 1 and 2, respectively; *m*=days pregnant; and *a, c, d, g* and h are the original parameters.

When regression analysis was performed on this function, it was found that the parameter h varied considerably from record to record and it was therefore decided to keep this figure constant. The parameter *h* was tested under many values and it was found that the most satisfactory value was *h =65*. Thus, the function to describe liveweight became:-

$$LW_n = \beta_0 + \beta_1(PC1) + \beta_2(PC2) + \beta_3(m-65)^3$$

where *PC1* and *PC2* = principal component scores 1 and 2, respectively; m=days pregnant, and β_1, β_2 and β_3 are regression parameters. It was found that the effect of multicollinearity is weak in this model and that the residuals are homoskedastic, independent and normally distributed (Table 2).

Table 2: *Goodness-of-fit and Analysis of Residuals of Liveweight Function*

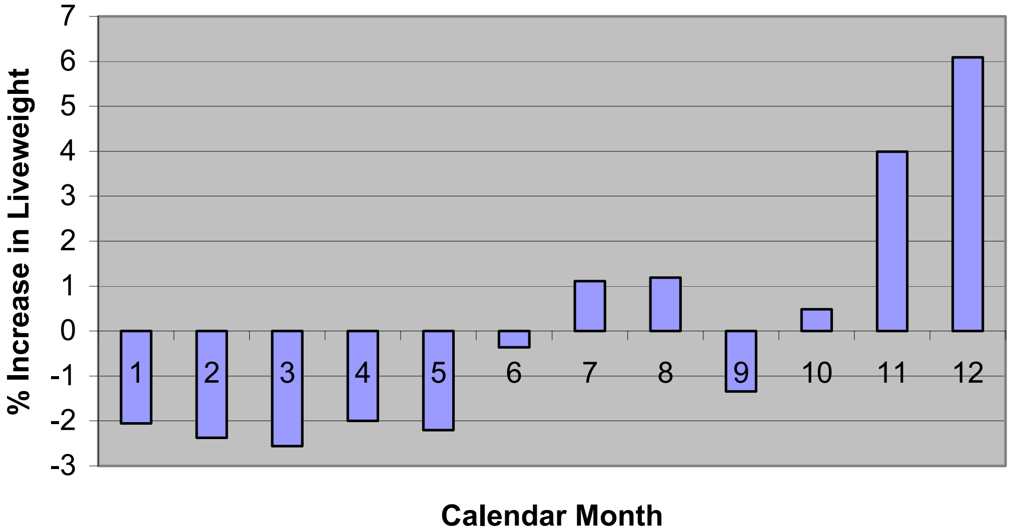
Test	New Function
MSPE	269.10
R^2	0.68
Autocorrelation	No
(Durbin-Watson)	2nd Order
Heteroskedasticity	No
(White's Test) [p-value]	0.36
Normality	Normal
(Kolmogorov-Smirnov)[p-value]	0.18
Kurtosis	0.80
Skewness	0.02
Multicollinearity	Weak
(Condition Index)	14.09

Finally the values for *a, c, d* and *g* were calculated using the values of β_1, β_2 and β_3 and the eigenvectors, α_{ij}, associated with the i^{th} variable and the j^{th} principal component. As lactation number was found to be significant the parameters of the new function were calculated separately for each of lactations 1, 2 and 3+ as shown below:

$$For\ Lactation\ \ 1: LW_n = 524.25 - 0.56\sqrt{n} - 1.53\ \ln(n) + 0.000033\ (m-65)^3$$
$$For\ Lactation\ \ 2: LW_n = 541.51 + 11.73\sqrt{n} - 1.75\ \ln(n) - 0.000030\ (m-65)^3$$
$$For\ Lactation\ \ 3+: LW_n = 594.15 + 2.60\sqrt{n} - 1.54\ \ln(n) - 0.0000034\ (m-65)^3$$

As the environment factors are known to have a significant effect on liveweight (Wood *et al,* 1980), figure 1 shows the incremental adjustment for environmental and seasonal effects on the liveweight model.

Figure 1: *Seasonal Deviations on the new function.*



The implication of these seasonal effects is that although the function can predict liveweight, it is also influenced by a seasonal component. Figure 1 shows that from January to May the liveweight function overestimates the liveweight by between 2 and 2.5%. This accounts for the variation of an early turn out to grass by some farmers and not by others. In the summer months, when the weather is less extreme, it can be seen that the environment has less of an influence on the predicted liveweight. In September there is a decrease in expected liveweight which is probably due to the fact that the grass is of poor quality at this time of year. In November and December the liveweight function underestimates the liveweight of cows and this is due to the effects of the environmental and nutritional factors and possibly conditions of animal housing.

Conclusions

The aim of this study was to arrive at a well-fitting and robust form of model to represent the shape of the liveweight curve for Irish dairy cows. An examination of the liveweight data using splines indicated that a four dimensional model was required. A number of models cited in the literature were also examined, these models being judged on their ability to adhere to the assumptions, which are made when fitting the models using regression analysis. The only assumption that was not satisfied was that of the explanatory variables being independent in every case. Thus, it was deduced that liveweight is in fact a function of age, lactation and pregnancy. It was evident, by examining the variance inflation factor values, that there was a strong correlation between two of the variables and as a result these were replaced by two linear independent components. Before using this model to predict the liveweight of a specific cow, adjustments are made to account for seasonal effects on liveweight. Thus the liveweight model is of the following form:-

$$\textit{For Lactation } 1: LW_n = 524.25 - 0.56\sqrt{n} - 1.53\ln(n) + 0.000033\,(m-65)^3 + \textit{seasonal effects}$$

$$\textit{For Lactation } 2: LW_n = 541.51 + 11.73\sqrt{n} - 1.75\ln(n) - 0.000030\,(m-65)^3 + \textit{seasonal effects}$$

$$\textit{For Lactation } 3+: LW_n = 594.15 + 2.60\sqrt{n} - 1.54\ln(n) - 0.0000034\,(m-65)^3 + \textit{seasonal effects}$$

In conclusion, the liveweight function, which accounts for the effect of age, lactation and pregnancy, is the best fitting model to explain the liveweight curve of Irish dairy cows; the effect of multicollinearity is weak and the residuals are normal, homoskedastic and independent.

References

Bakker, H., Koops W.J. 1978. An approach to the comparison of growth curves of Dutch Friesian, British Friesian and Holstein Friesian cows. Current Topics in Veterinary Medicine. Pattern of Growth and Development in Cattle 2, 705-715

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

Berglund, B., Danell, B. 1987. Live weight changes, feed consumption, milk yield and energy balance in dairy cattle during the first period of lactation. Acta Agric. Scand. 37, 495-509.

Berry, D.P., Horan, B., Dillon, P. 2004. Comparison of growth curves of three strains of female dairy cattle. Anim. Sci. (In Press)

Brown, J.E., Fitzhugh, H.A., Jr and Cartwright, T.C. 1976. A comparison of nonlinear models for describing weight-age relationships in cattle. J. Anim. Sci. 42, 810-818

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

Forbes, J.M. 1983. Models for the prediction of food intake and energy balance in dairy cows. Livest. Prod. Sci. 10, 149-157

Gravir, K. 1967. Studies on Different Body Measurements as Estimators of live and carcass weight in young NRF bulls. Acta Agric. Scand. 17, 217-227

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.

Heinrichs, A.J., Rogers, G.W., Cooper, J.B. Predicting Body weight and wither height in Holstein heifers using body measurements. J. Dairy Sci. 75, 3576-3581

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

Korver, S., Van Arendonk, J.A.M., Koops, W.J. 1985 A function for live-weight change between two calvings in dairy cattle. Anim. Prod. 40, 233-241.

Lopez-Villalobos, N., Lemus-Ramirez, V., Holmes, C.W., Garrick, D.J. 2001. Lactation curves for milk traits, live weight and body condition score for heavy and light Holstein-Friesian cows. Proc. New Zealand Soc. Anim. Prod. 61, 217-220

Madalena, F.E., Teodoro, R.L., Madureira, A.P. 2003. Relationships of weight and height with age in hybrid Holstein Friesian /Guzera females. Genet. Mol. Res. 2, 271-278

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

Moore, A.J. 1985. A mathematical equation for animal growth from embryo to adult. Anim. Prod. 40, 441-453

Perotto, D., Cue, R.I., Lee, A.J. 1992. Comparison of nonlinear functions for describing the growth curve of three genotypes of dairy cattle. Can. J. Anim. Sci. 72, 773-782

Quinn N., Killen L., Buckley F. 2004. Empirical Algebraic Modelling of Lactation Curves using Irish Data. J. Ir. Agric. & Food Res. (Submitted)

Taylor, St. C.S. 1980. Genetically Standardized Growth Equations. Anim. Prod. 30, 167-175

Walter, J.P., Mao, I.L., Emery, R. S. 1984. Simultaneous fitting of energy requirements and intake, lactation response and body weight changes in dairy cattle, with different forage bases rations. J. Dairy Sci. 70, Supplement, Abstract P221.

Wicks, H.C.F. 2001. The influence of genetic merit and environment on dairy cattle performance on farm. PhD Thesis, Imperial College of Wye, University of London

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., King, J.O.L., Youdan, P.G. 1980. Relationships between size, liveweight change and milk production characters in early lactation in dairy cattle. Anim. Prod. 31, 143-151