# Modelling Lactation and Liveweight Curves in Irish Dairy Cows

Noreen Quinn

B Ed , M Sc

A thesis submitted in fulfillment of the

requirements for award of

Doctor of Philosophy (Ph D )

to the

Dublin City University

Faculty of Engineering and Computing

School of Computing

Supervisor Dr Lynn Killen

July, 2005

©Noreen Quinn 2005

# DECLARATION

*I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work*

| | |
|---|---|
| Signed | (Noreen Quinn) |
| Student ID | 50219774 |
| Date | $2^{nd}$ September 2005 |

# Acknowledgements

A journey is easier when you travel together (as I have learned from the weekly train journey to Cork) and this thesis is the result of many contributions from numerous sources to whom I am indebted for their constant support, patience and inspiration

Pre-eminent among them is Dr Seamus Crosse of Teagasc without whose support and enthusiasm, this project would never have been initiated He secured funding for the first year of this study and subsequently a Teagasc Walsh Fellowship for the remaining time was granted Without this financial support from Teagasc this study could not have been undertaken

I would especially like to thank Dr Frank Buckley, my Teagasc supervisor As I didn't grow up on a dairy farm, Frank ensured that the outcomes from my research were agriculturally plausible I also wish to acknowledge Mr Dave Cliffe, Dr Laurence Shalloo, Dr Ross Evans and the others at Moorepark, Teagasc for supplying the data and for giving some advice

I would like to thank my supervisor Dr Lynn Killen From the time Lynn accepted me as a student she has guided me extremely well With her rigorous assessment and support she has assisted me in converting an idea into a scientifically acceptable piece of work She encouraged me to write regularly and to submit to numerous conferences and journals, which have helped me to get my work to this level I acknowledge Lynn's approachable character and the fact that her door was always open to discuss problems Thank you so much

Many thanks are also extended to Mr Gary Keogh, Ms Patricia Gunning and Mr Adel Sharkasi of DCU's Modelling and Scientific Computing group for the advice received and to Dr Victor Olori and Mr Martin Burke from the Irish Cattle Breeding Federation I would also like to thank my other fellow postgraduate students who helped to make this journey easier Karl, Noel, Niall, John, Puspita, Ana, Ronan, Colm, Grainne, Tommy, George and Claire (You can have your LaTeX book back now) to name but a few! To Maria, Fiona, Suzanne, Michelle and Maria - Thank you!

To my parents, Helen and Harry and my brothers Anthony, Raymond and Kieran, I extend my love and gratitude They have always supported me through the ups and downs and in no small part do I owe the success of this achievement to them

Finally, to John, thanks for showing me the more practical side of the dairy industry and whose patience, understanding and love never ceased to astound me I love you

Thank you all, so much

*Noreen Quinn, 02/09/2005*

Dedicated to

*my parents, Harry and Helen*

They began my education

They motivated me to continue it

They will always contribute to it

and to

*John*

who managed to take my mind off work every weekend

# CONTENTS

# List of Figures

# List of Tables

# Modelling Lactation and Liveweight Curves in Irish Dairy Cows

**Noreen Quinn**

Dublin City University,

Glasnevin,

Dublin 9

## ABSTRACT

The purpose of this research has been to model the milk yield, fat content, protein content and liveweight of Irish dairy cows over a lactation period The analysis was carried out on 15,729 lactation records from commercial and experimental herds including both autumn and spring calving animals of various breeds Initially, the factors which affect milk yield, its constituents and liveweight were investigated and then a scientific method for detecting abnormal recordings was formulated This method of detection is very effective and abides by the guidelines outlined by the International Committee of Animal Recording After removing the abnormal recordings from the data, a number of models were fitted and their suitability was assessed on the basis of their goodness-of-fit and adherence to the assumptions made in carrying out regression analysis When modelling milk yield, a severe problem of multicollinearity was encountered and methods of reducing multicollinearity were investigated As a result a new model was developed, which provided an acceptable level of accuracy in representing the shape of the lactation curve for Irish dairy cows A model that satisfied the assumptions of regression analysis and predicted the actual content of the constituents to within 0 01 per cent of the actual values was also developed while a novel approach was used for modelling liveweight Firstly, splines were used to find the dimensions of the data and principal component regression

was used to estimate the regression coefficients of this new model This model satisfactorily represented the shape of the liveweight curve and it can be easily updated for use by bio-economists The models proposed in this thesis are currently being incorporated into the Moorepark Dairy Systems Model which investigates the challenges that currently face the Irish dairy industry

# CHAPTER 1

# INTRODUCTION

"Denmark has a smaller national milk quota than Ireland yet produces three times as much cheese - 300,000 tonnes"

"Yield per cow in Ireland is 60 per cent of that achieved in Holland or Denmark"

These are two quotes from the Chief Executive Officer of Glanbia (Ireland's largest manufacturer of dairy products), John Maloney, at the National Dairy Conference in 2002 These two statements ask one major question "Why is this the situation?" The simple answer is obviously related to efficiency but the problem is how does one address this?

## 1 1 The Irish Dairy Industry

In the coming years the Irish dairy industry will be confronted by many challenges and unless it can respond strategically and effectively the industry will be seriously undermined Traditionally the Irish dairy industry's emphasis has been on commodity products such as butter and skim milk powder which have been supported by the European Union (EU) through export subsidies, intervention or internal market supports  Ireland still has a high reliance on these products, as can be seen in Figure 1 1 (Promar International, 2003), while other European countries have altered their product portfolio mix  However, there is a strategic plan in place to alter Ireland's product mix and by 2015 five per cent of its product portfolio mix should be in higher-value functional and organic foods with a reduction of 20 per cent in commodity based products  The underlying quality of milk being supplied to production plants has an effect on the type of product that can be produced, and it is essential to have year round supplies of good quality milk in order to serve European markets for quality fresh products (Maloney, 2002)

The productivity issue referred to in the second quote can also be improved through improved efficiency  Studies show that if producers milk fewer cows for longer, they can increase their total milk yield per cow and can deliver their quota in a more even supply pattern throughout the year and at lower cost than is currently the case (McCarthy, 2000)  At present, Irish milk producers, who are supplying milk for processing into dairy products, are adjusting the date of calving so that the cows calve at the time of lowest milk production cost (Table 1 1), thus maximising production cost efficiency from a grass-based production perspective  This calving pattern results in high levels of supply in the peak milk supply months of March to June

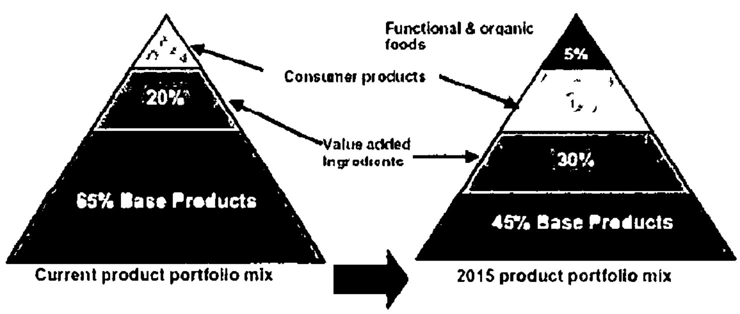


Figure 1 1  Dairy product portfolio 2003-2015

Table 1 1 Percentage of the national herd calving per month

| Month | 1991 | 1998 | 2000 | 2001 |
|---|---|---|---|---|
| January | 11 9 | 10 1 | 7 6 | 5 9 |
| February | 20 3 | 20 8 | 17 5 | 17 3 |
| March | 24 3 | 25 9 | 23 5 | 24 4 |
| April | 16 4 | 17 4 | 19 1 | 19 9 |
| May | 8 8 | 9 8 | 12 0 | 11 5 |
| June | 4 2 | 4 1 | 5 9 | 5 8 |
| July | 2 4 | 2 3 | 3 0 | 3 3 |
| August | 1 8 | 1 6 | 2 0 | 2 3 |
| September | 2 0 | 1 8 | 2 3 | 2 5 |
| October | 2 4 | 2 1 | 2 4 | 2 6 |
| November | 2 3 | 2 0 | 2 3 | 2 3 |
| December | 3 2 | 2 2 | 2 3 | 2 4 |

*Source* Department of Agriculture and Food (2002)

Table 1 2 illustrates how the seasonality of milk supply in Ireland has gradually disimproved over the years In 1999 the peak month production (May), as measured by milk deliveries, was 5 6 times the lowest month's production (January) This ratio has gradually disimproved over the last number of years, having been as low as 4 7 in 1993 (Promar International, 2003) In Northern Ireland the pattern of milk supply is much closer to that experienced throughout the rest of Europe, with a peak to trough ratio of 1 6, which demonstrates that it is possible to improve seasonality significantly in Ireland To handle the peak milk supply the average capacity utilisation of milk processing plants in Ireland is close to 60 per cent In Denmark and the Netherlands, who are Ireland's main EU competitors, the excess

Table 1 2 Summary of milk supply

| Milk Supply ('000 tonnes) | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
|---|---|---|---|---|---|---|
| Peak month milk deliveries | 740 | 748 | 700 | 710 | 717 | 731 |
| Trough month milk deliveries | 137 | 131 | 124 | 119 | 122 | 122 |
| Peak to trough ratio | 5 4 | 5 7 | 5 6 | 6 0 | 5 9 | 6 0 |

*Source* Department of Agriculture and Food (2002) processing capacity is only three per cent, meaning that all processing assets are used throughout the year In Ireland the processing plants are severely under-utilised in the winter period While Ireland's lower capacity utilisation enables processors to switch between products in response to short-term market trends, it tends to put Ireland at a disadvantage in the European dairy market If the problem of seasonality of milk supply is addressed it will increase capacity utilisation in the production plants and also deliver more opportunities for the industry to add further value to Irish dairy products To address the issue of the seasonality of milk supply, the national calving pattern needs to be changed so as to produce a more even supply of milk over the year In order to assess the impact any change in calving pattern would have on the seasonal pattern of milk supply, a thorough understanding of lactation curves is necessary

## 1 2 The Importance and Relevance of Lactation Curves

A lactation curve is a summary of the longitudinal milk yield of a cow from calving until drying off prior to a subsequent calvmg From this cumulative lactation curves can be estimated and total lactation milk yields may be predicted from incomplete data Appropriate models provide the basis for examining the effect that changing the calvmg pattern will have on the milk supply pattern A mathematical model of the lactation curve provides summary information about dairy cattle production, which is useful in making management and breeding decisions as well as being useful in simulating a dairy enterprise Generally, the objective in modelling the lactation curve is to predict the yield on each day of lactation with minimum error so as to elucidate the underlying pattern of milk production in the presence of high local variation due to the effect of the environment A good model should be capable of predicting total yield from a few test day records and thus minimises the cost of milk recording in Ireland Such a model is applicable for dairy producers and processors and could be invaluable in the event of natural phenomena (such as Foot and Mouth

Disease) which restricts milk recording, without unduly jeopardising accuracy of genetic evaluation for the farmer However, the usefulness of any mathematical model depends on how well it can simulate the biological process of milk production and adjust for factors affecting it (Olori et al , 1999)

Milk production in a dairy cow typically rises to a peak in the first 40 to 70 days post partum and declines thereafter (Wood, 1967) Several models have been developed to describe such a lactation pattern which is assumed to be the same for all cows (Wood, 1967, Yadav et al , 1977, Cobby and Le Du, 1978, Keown and Van Vleck, 1973, Ali and Schaeffer, 1987, Wilmink, 1987, Elston et al , 1989, Perochon et al , 1996) The goodness-of-fit of a lactation curve model may depend on whether the objective is to predict the cumulative yield or individual daily yields In the former case the residuals of the predicted yields may not be important as long as they sum to zero while in the latter case the magnitude and distribution of the residuals are important It is also necessary to distinguish between modelling the mean lactation curve of a group of cows with different lactation curves and modelling individual lactations In addition, differences between various breeds of cow, in relation to lactation curves, is of particular interest at the present time in Ireland and will be investigated for the first time in this study

Some studies on estimating the shape of lactation curves have been undertaken in Ireland, but it has been quite some time since these were reviewed The most recent review was by Crosse et al (1988) and previously to that by Killen and Keane (1978) As conditions in Irish agriculture have changed considerably since then it is appropriate, and indeed necessary, that this research is reviewed, re-analysed, and improved upon

Modelling total lactation of milk yield and its constituents for dairy cows is the basis of this study Initially the factors which affect lactation yield will be investigated and models will then be proposed to model the milk yield, fat content and protein content of milk One element which will be examined for the first time in Ireland, is the variation in the shape of the lactation curve between breeds and its variation with season In addition the liveweight of dairy cows over a lactation period will be investigated, also for the first time in Ireland

## 1 3 The Scope of the Thesis

There are three main sections to this thesis Each of these sections involves evaluating the models which are already cited in literature, judging the suitability of these models to Irish data and exploration of new forms of models As a result of finding a robust model, which is relatively simple to estimate and use, a seasonal production pattern table can be created for use by bio-economists The three sections to this study are

1 Modelling the lactation curve of milk yield

2 Modelling the fat and protein content curves

3 Modelling liveweight

After reviewing the literature (Chapter 2) it was found that the use of empirical algebraic models was the preferred methodology for modelling the milk yield, fat and protein content and liveweight curves of Irish dairy cows It suggests that linear and nonlinear regression are the necessary forms of modelling required, and that the assumptions of regression analysis need to be examined for each model so as to arrive at a robust and well-fitting model

Assessing and examining the factors which influence milk yield, fat content and protein content provides valuable information to the dairy farmer The success of a dairy farmer really depends on how well they can predict the lactation curves of their cows (Wood, 1969) Chapter 3 examines the factors which affect the level of milk production of Irish dairy cows, of different breeds, for the first time since the extensive advances in Irish agriculture over the last twenty years It also analyses these factors and demonstrates the effects they have on milk yield and its constituents

As abnormal recordings will inevitably occur in large datasets of the type being analysed in this study, Chapter 4 illustrates a method which is effective in detecting abnormal recordings in Irish milk recorded data This method abides by the guidelines outlined by the International Committee of Animal Recording (ICAR, 2002) and is an appropriate method for detecting abnormal recordings In addition, it could be particularly useful with the introduction of electronic milk recording devices in the near future

The purpose of Chapter 5 is to find a well-fitting, robust, single equation model which describes the shape of the milk yield curve for Irish dairy cows This chapter examines the suitability of a number of algebraic models cited in the literature, using Irish data The models have been evaluated using a novel approach, by examining their goodness-of-fit and their adherence to the assumptions of regression analysis and if deemed necessary the models have been modified Much of the material presented in this chapter has been published in the proceedings of CompStat 2004 (Quinn et al , 2004) and ICAR[1] 2004 (Quinn et al , 2005b) and is due to appear in the Irish Journal of Agricultural and Food Research Quinn et al (2005a)

The fat and protein content of milk is modelled in Chapter 6 Many models which were cited in the literature, to model milk yield, were also examined on their suitability to model the curves for these constituents It was found that over a lactation period one of these models was consistent in predicting weekly fat and protein content and was also reasonably accurate in predicting the average fat and protein content

The liveweight change for Irish dairy cows over a lactation period is modelled in Chapter 7 As there are very few models to describe the pattern of liveweight change over a lactation in the literature, various approaches were considered Initially, time series techniques were examined as the data is inherently of a time series nature Splines were also examined so as the dimensions of the model required to

[1] International Committee of Animal Recording

fit the data could be approximated However, as an incomplete gamma function, which was previously used to model milk yield, has been used in other studies to model liveweight, other milk yield models were investigated It was a result of these investigations, that an equation was derived to model liveweight change over a lactation period

Finally, Chapter 8 provides a summary of the thesis incorporating the conclusions and implications

# CHAPTER 2

# LITERATURE REVIEW

## 2 1 Introduction

This review focuses initially on the different approaches used to model milk yield and its constituents over a lactation period The concentration of either fat or protein in the milk produced throughout a lactation can be represented by a curve whose shape will normally "mirror milk yield" (Pulina, 1990), as milk yield increases, the fat and protein content decrease and vice versa Many researchers who have investigated the shape of the milk yield curve have also investigated the shape of the fat and protein curves Therefore in this review, the literature related to the fat and protein fractions of milk will be considered together with the literature associated with milk yield modelling

Liveweight and its changes have been found to be important, in recent years, for the formulation of diets (Korver et al , 1985), for the calculation of dry matter intake prediction formulae (Devir et al , 1995) and for inclusion in an economic model to simulate the dairy production system (Shalloo et al , 2004) The second section of this chapter will discuss the different approaches to modelling liveweight which are cited in the literature and highlight the approach which is investigated later in this study

## 2 2 Modelling Milk Yield, Fat and Protein

There are many approaches to modelling milk yield and its constituents Initially, studies focused on two methodologies empirical regression models which used regression analysis to find estimates of the milk yields in each week of lactation, and test day modelling which examined the relationship between test day yields and several explanatory variables In more recent years lactation curves have been examined as multiphasic functions whereby the lactation curve is segmented into different phases and an equation is fitted separately to each phase, or by using Bayesian analysis which endeavors to estimate parameters of an underlying distribution based on the observed distribution Test-day modelling procedures have been developed and subsequently divided into sub-categories namely repeatability test day models, random regression models and reduced rank models Another approach for modelling milk yield is by using autoregressive procedures which distinguish between the environmental effects due to the cow within and between lactations

### 2 2 1 Empirical Algebraic Models using Regression

Since early in the $20^{th}$ century, empirical regression models of the shape of the milk yield curve have been proposed by authors such as Brody et al (1923), Brody et al (1924), Sikka (1950), Dave (1971), Wood (1967), Wilmink (1987), Ali and Schaeffer (1987) and Guo and Swalve (1995)

Brody et al (1923) proposed the following equation

$$Y_n = ae^{-kn} \qquad (2\ 2\ 1)$$

where $Y_n$ is the milk yield in lactation week $n$ and there is a constant relative rate of decline in yield of $k$ kg/kg per week from an initial value of $a$ This equation was generally used to model only the declining phases of lactations, but it has been used to model whole lactations by a number of authors including Singh and Bhat (1978) and Mukundan and Bhat (1983) Brody et al (1924) proposed a more elaborate equation that was derived from the difference of two exponential functions This equation was a model for the whole lactation and it took the following form

$$Y_n = ae^{-bn} - ae^{-cn} \tag{2 2 2}$$

where $Y_n$ is the milk yield in lactation week $n$ and $a, b$ and $c$ are the parameters The main problem arising with this equation is that it predicts an initial yield of zero which is unrealistic It was found by Cobby and Le Du (1978) that this model underestimated yields in mid lactation and overestimated them in late lactation

Wood (1967) was the first to create an equation that represented the lactation curve reasonably accurately His model, which has since become known as the incomplete gamma function, is still being used to predict milk yield today and takes the following form -

$$Y_n = an^b e^{-cn} \tag{2 2 3}$$

where $Y_n$ is the yield in week $n$, $a$ is a scaling factor associated with the average yield, $b$ is related to pre-peak curvature and $c$ is related to post-peak curvature This model uses the method of least squares to obtain estimates for three regression parameters $a, b$ and $c$ If $c \ll 1$, then the value of the parameter $a$ is approximately equal to the yield immediately after calving Wood's model predicts a peak yield of $a(\frac{b}{c})^b e^{-b}$ which occurs $\frac{b}{c}$ weeks after calving As Wood's equation is intrinsically non-linear and it was difficult, at the time, to perform nonlinear regression, many researchers transformed it into linear form by taking the natural logarithm of both sides of the equation as follows

$$\ln(Y_n) = \ln(a) + b\ln(n) - cn \tag{2 2 4}$$

Another alternative to nonlinear least squares estimation of the parameters in the model of Wood (1967) was to use a weighted multiple regression of $ln$ $(Y_n)$ on $ln$ $(n)$ and $n$ This provided an approximation to the nonlinear procedure by means of a simple weighted linear least squares analysis Cobby and Le Du (1978) showed that the appropriate weights are proportional to $Y_n^2$, i e the square of the corresponding yield Wood (1969) examined the factors which affect the shape of the curve and from this he developed a technique whereby it is possible to predict a cow's milk production from month to month Wood (1976) used this same model to describe the production and concentration of both fat and protein It has been found, however, that the fit of Wood's curve to monthly data can be poor, especially in subtropical and tropical climates (Kellogg et al , 1977, Shanks et al , 1981) and therefore alternative models have been proposed

Yadav et al (1977) approached the problem of fitting an equation to milk yield data and arrived at an inverse quadratic polynomial of the following form

$$Y_n = \frac{n}{a + bn + cn^2} \tag{2 2 5}$$

The variable $Y_n$ is the weekly milk yield in the $n^{th}$ week of lactation, $a$ and $c$ are the constants which describe the rising and declining extremes of the lactation curve respectively and $b$ is the average slope of the curve Kumar and Bhat (1979) found that this model gave a good fit for lactations which start at a low level and peak earlier than average Later studies of empirical regression models, however, found this model to be less satisfactory (Papajcsik and Bodero, 1988, Olori et al , 1999)

An exponential plus linear decline model was proposed by Cobby and Le Du (1978) as a follow-up to the model of Wood (1967) They replaced the term $n^b$ in the model of Wood (1967) by the asymptotic curve $1 - e^{-qn}$, resulting in the following

$$Y_n = ae^{-kn} - ae^{-dn} \tag{2 2 6}$$

For large values of $n$ (post-peak) the milk yield approximately follows an exponential decline (Gaines, 1927) of $k$ kg/kg per week Replacing the first exponential in equation (2 2 6) by the line $a - kn$ gives the curve

$$Y_n = a - bn - ae^{-cn} \tag{2 2 7}$$

where $Y_n$ is the milk yield in lactation week $n$ and $a, b$ and $c$ are the parameters This model predicts peak milk yield to occur at $c^{-1}ln(\frac{ac}{b})$ weeks after calving, with an approximately linear ($Y_n = a - kn$) decline thereafter (Fischer, 1958) where $k$ is now measured in kg/week and $\frac{a}{k}$ is an estimate of the length of lactation

In more recent times, Wilmink (1987), Ali and Schaeffer (1987) and Guo and Swalve (1995) created models which, they claimed, better represented the shape of the lactation curve Wilmink (1987) proposed the following non-linear parametric curve with four parameters, to predict the milk, fat and protein yields from cows

$$Y_n = a + bn + ce^{-dn} \tag{2 2 8}$$

where $n$ is days in milk i e days since calving As days in milk increases, the exponential term, which is associated with the yield in the early stages of lactation, tends to zero, and the decline in yield after the peak is eventually represented by the straight line $a + bn$ Usually the fourth parameter, $d$, is held constant, reducing the number of parameters to be estimated from four to three and greatly simplifying the fitting of the curve Brotherstone et al (2000) showed $d$ to be consistent over lactations and over age groups within lactation for UK data, at a value of 0 10, while Olori et al (1999) estimated $d$, following a preliminary analysis, to be 0 61 for their dataset

While the model of Wilmink (1987) was being developed in the Netherlands, the model of Ali and Schaeffer (1987), which is based on polynomial regression, was being developed in Canada The regression model proposed by Ali and Schaeffer

(1987) was of the following form

$$Y_n = a + b\gamma + c\gamma^2 + d\omega + e\omega^2 + f \tag{2 2 9}$$

where $\gamma = \frac{n}{305}$ , $\omega = ln(\frac{305}{n})$ and $n$ = days since calving or days in milk, $a, b, c, d$ and $e$ are the regression co-efficients where $a$ is associated with the peak yield, $d$ and $e$ with the increasing slope of the curve and $b$ and $c$ are associated with the decreasing slope, $f$ is the residual error for this model This regression model requires at least six test weighings in order to estimate the parameters, which is a disadvantage for some applications, especially when using this model for extending part lactation records Ali and Schaeffer (1987) used this model specifically for predicting milk yield, because it is a polynomial regression model it keeps its concave shape and therefore could not be used to model the fat and protein content of milk

Guo and Swalve (1995) proposed the mixed logarithmic model

$$Y_n = a + b\sqrt{n} + c\log(n) \tag{2 2 10}$$

where $n$ is the number of weeks since calving Olori et al (1999) examined this model along with those of Wood (1967), Wilmink (1987) and Ali and Schaeffer (1987) and found that the combined exponential and linear model proposed by Wilmink (1987) was the best three-parameter model for predicting herd mean yield, having both the smallest and least correlated residuals It also performed adequately in fitting individual lactation data

### 2 2 2 Test-day Models and their Sub-categories

Other studies, have dealt specifically with the shape of the lactation curve by relating individual test day yields to the stage of lactation at which they were recorded Keown and Van Vleck (1973) proposed a model

$$Y_{ijk} = \mu + (HY)_i + (SAS)_j + \epsilon_{ijk} \tag{2 2 11}$$

where $Y_{ijk}$ is the milk (or fat) production on an individual test day $k$ in the $i^{th}$ herd-year and the $j^{th}$ season-age-stage, $\mu$ is an unknown constant, $(HY)_i$ is the effect of the $i^{th}$ herd-year in their dataset and $(SAS)_j$ is the effect due to the $j^{th}$ season-age-stage This approach allows a residual variance-covariance matrix to be defined such that estimates of yield for missing test days can be based on their estimated covariances with observed test days This type of approach has also been examined in other studies such as those of Schaeffer and Burnside (1976), Stanton et al (1992), Ptak and Schaeffer (1993) and Lee et al (1995) However, significant developments in the use of test day models have occurred and therefore this area of modelling lactation curves has branched into three sub-categories repeatability test day models (Reents et al , 1998, Ptak and Schaeffer, 1993), random regression test day models (Schaeffer et al , 2000, Schaeffer and Dekkers, 1994, Jamrozik and Schaeffer, 1997, Jamrozik et al , 1997, Brotherstone et al , 2000, Kettunen et al , 2000) and reduced rank test day models (Lidauer et al , 2000, Anderson, 2002)

The repeatability test day model considers observations within lactation as repeated observations (Ptak and Schaeffer, 1993, Reents et al , 1995) but the stage of lactation is considered only in the fixed effect part of the model It assumes that there is a standard shape to the lactation curve for all cows in the same age-season subclass, and that the estimated additive genetic effects of animals are reflected in the differences in the height of these curves Differences in persistency (the length of time the cow maintains peak production during a lactation) are ignored in this method

Schaeffer and Dekkers (1994) proposed an extension of the repeatability test day model by allowing the shape of the lactation curve to differ for individual cows by including random regression coefficients for each animal Two sets of regressions were involved fixed regressions for all cows belonging to the same subclass of age-season of calving to describe the shape for that cow, and the random regression for a cow to describe the deviations from the fixed regressions This random regression allows cows to have differently shaped lactation curves Jamrozik and Schaeffer (1997) and Jamrozik et al (1997) examined random regression models and concluded that further study on appropriate functions that can be used in a random regression model seemed warranted, as differences between the models that were chosen in their study may not be that important on a practical basis Brotherstone et al (2000) investigated other functions suitable for the analysis of daily milk yield, using random regression models, their results showed that the measurement error variances were generally lowest around peak lactation, and higher at the beginning and the end of lactation Kettunen et al (2000) suggested that due to the statistical complexity of random regression models, the use of multitrait models would be more useful for modelling the lactation curves of dairy cows

The third sub-category of test-day models is the reduced rank model Lidauer et al (2000) used the following multiple-trait multi-lactation reduced rank random regression test-day model

$$
\begin{aligned}
\begin{bmatrix} y_{Fijklmnoq} \\ y_{Lijklmnopq} \end{bmatrix} = {} & \begin{bmatrix} age_{Fi} \\ age_{Li} \end{bmatrix} + \begin{bmatrix} dcc_{Fj} \\ dcc_{Lj} \end{bmatrix} + \begin{bmatrix} (ym)_{Fkl} \\ (ym)_{Lkl} \end{bmatrix} \\
& + \begin{bmatrix} \phi(DIM)\mathbf{b}_{Fm} \\ \phi(DIM)\mathbf{b}_{Lm} \end{bmatrix} + \begin{bmatrix} (hy)_{Fnk} \\ (hy)_{Lnk} \end{bmatrix} + \begin{bmatrix} (htm)_{Fnkl} \\ (htm)_{Lnkl} \end{bmatrix} \\
& + \begin{bmatrix} \sum_{r=1}^{6} s(DIM)_{Fr} a_{or} \\ \sum_{r=1}^{6} s(DIM)_{Lr} a_{o(r+6)} \end{bmatrix} + \begin{bmatrix} \sum_{r=1}^{6} t(DIM)_{Fr} p_{or} \\ \sum_{r=1}^{6} t(DIM)_{Lr} p_{o(r+6)} \end{bmatrix} \\
& + \begin{bmatrix} \\ \sum_{r=1}^{6} t(DIM)_{Lr} w_{op(r)} \end{bmatrix} + \begin{bmatrix} e_{Fijklmnoq} \\ e_{Lijklmnopq} \end{bmatrix}
\end{aligned}
$$

where $y_{Fijklmnoq}$ are the first lactation test-day observations of milk, protein, and fat yields, and $y_{Lijklmnopq}$ are later lactation observations The fixed effects are age at calving ($age$), days carried calf ($dcc$), test-year x test-month ($ym$), regression coefficients or the shape of the lactation curve ($(\phi(DIM)\mathbf{b})$), and the herd-year ($hy$) of the test The stage of lactation is modelled by an interaction of lactation curve, calving season, calving year and lactation number where there are three calving seasons, November - February, March - June, and July - October and three categories of lactation number $(1, 2, 3+)$ The covariables $\phi(DIM) = [c_1 c_2 c_3 c_4 c_5]$, where $c_1, c_2$, and $c_3$ represent a quadratic Legendre polynomial by days in milk ($DIM$), and $c_4$ and $c_5$ are exponential terms $exp(-p_1 DIM)$ and $exp(-p_2 DIM)$, where $p_1$ is 0 05 for milk yield and 0 10 for the other traits, and $p_2$ is $0\ 06, 0\ 01$, and 0 35 for milk, protein and fat yield respectively of first lactation cows and $0\ 04, 0\ 20$, and 0 35 respectively for milk, protein and fat yield of second and later lactation cows Although reduced rank models reduce the number of coefficients to be estimated per animal, if the model is misspecified it causes the effects of under- and over- estimating of the rank of the reduced form, which leads to biases occurring or unnecessarily large variances of the estimators and predictors (Anderson, 2002)

### 2 2 3 Multiphasic Approach

A multiphasic approach to modelling lactation curves was considered by Grossman and Koops (1988) The multiphasic function suggests that the lactation of a cow is made up of several phases, and by cumulating the production in each of these phases, the function estimates the total milk production The main advantage of this approach is that it creates smaller, more random residuals than the incomplete gamma function (Wood, 1967), it predicts the total 305-day yield without approximation and its parameters are easily interpretable and have biological importance (De Boer et al , 1989) Grossman and Koops (1988) concluded that two lactation phases (pre- and post- peak) were sufficient to model the lactation curve of milk yield However, De Boer et al (1989) found that a triphasic function was necessary to model milk yield while fat yield could be modelled using a diphasic model Sherchand et al (1995) and Vargas et al (2000) also examined multiphasic models, specifically multiphasic logistic models These models were of a form similar to that of Grossman and Koops (1988), as follows

$$Y_n = \sum_{i=1}^{p}(a_i b_i[1 - tanh^2(b_i(n - c_i))]) \qquad (2\ 2\ 12)$$

where $Y_n$ is the milk yield on day $n$, $p$ is the number of lactation phases, $tanh$ is the hyperbolic tangent function, $a_i b_i$ is the peak yield for phase $i$, $c_i$ is the peak day for phase $i$ and $2b_i^{-1}$ is the time taken (in days) to attain about 75 per cent asymptotic yield during phase $i$ It was found that multiphasic logistic models are intrinsically suitable to describe extended lactations (Vargas et al , 2000) but that a computer must consistently monitor the daily yield of individual cows before using this approach (Sherchand et al , 1995) The major criticism of this approach is that there appears to be no physiological foundation for assuming that lactation is a multiphased process (Tozer and Huffaker, 1999), even though the function appears to behave well and provide a valid statistical result (Rook et al , 1993)

### 2 2 4 Bayesian Approach

Jones (1997) examined a Bayesian approach to the use of the model of Wood (1967) and found that it had many advantages The Bayesian approach was more accurate in predicting the milk yield and it was readily adapted to individual herds, however, Jones (1997) concluded that further investigation was necessary Rekaya et al (2000) and Jamrozik et al (2001) examined the models of Wood (1967) and Wilmink (1987) using Bayesian analysis, and concluded that in the future this method might be used, but that computing limitations prevented these methods from being used for routine national genetic evaluations The Bayesian approach to modelling lactation curves is very subjective, if the priors are not chosen correctly it can have a major consequence on the model being investigated

### 2 2 5 Autoregressive Procedures

Another approach which has been used in more recent times to model lactation curves is the autoregressive test day model (Macciotta et al , 2002, Vasconcelos et al , 2002) Although this procedure requires that the data is recorded at equal intervals, which is not the case with the dataset in this study, Carvalheira et al (1998, 2001) indicated that autoregressive test day repeatability models could estimate accurate lactation curves and predict all test day residual yields from all cows present in the analysis The autoregressive model used by Vasconcelos et al (2002) for modelling milk, fat and protein yields, was as follows

$$y_{ijkLmn} = HTD_i + Age(H)_j + DIM(H)_{k(L)} + p_{m(L)} + t_{n(mL)} + \epsilon_{ijkLmn} \quad (2\ 2\ 13)$$

where $y$ is the test day yield, $HTD$ is the fixed effect of herd-test-date, $Age(H)$ is the fixed effect of age at calving nested within herd, $DIM(H)$ is the fixed effect of days in milk (DIM) nested within herd and lactation $(L)$, $p$ is the random effect of the long term environmental effects accounting for the correlations generated by the cow across the lactation $(L)$, $t$ is the random effect of the short term environmental effects accounting for the correlations due to the cow between test day and within lactation, and $\epsilon$ is the random residual effect Both $p$ and $t$ are fitted with first order autocorrelation structures It was found that autoregressive models are simple to use and can be easily implemented in data recording software, even at farm level (Macciotta et al , 2002) However, they represent a relatively new approach to modelling lactation curves and need to be investigated further

### 2 2 6 Standard Lactation Curve Method in Current Use in Ireland

In Ireland, the SLAC (Standard Lactation Curve) method of Olori and Galesloot (1999) is the preferred method for predicting milk, fat and protein yield This method incorporates 2,160 lactation curves for milk and its constituents, accounting for variation in the effect of season, calving age and level of production It is acknowledged however, that having a library of equations, from which the most appropriate one is chosen, will almost inevitably give accurate predictions This method involves interpolation and initially the yields for a set of 15 fixed days are predicted Yields on the days between measured yields are obtained by linear interpolation while a 305-day yield is calculated for each lactation Yields on days before the first recording and after the last recording are obtained by fitting the model of Wilmink (1987) to the available test day records

### 2 2 7 Milk Yield and Milk Constituents Literature Summary

A wide variety of function forms have been suggested to predict the test day values of milk yield, fat content and protein content from a lengthy search for a robust model A model that performs well statistically over a wide variety of datasets, and that corrects minor biological defects perceived in previously applied models is what is required (Tozer and Huffaker, 1999) Most of the methods outlined above are based on parametric lactation curves of homogeneous groups of cows with information on individual cow variations Jones (1997) proposed an empirical Bayesian method in which milk yields from a lactation in progress are combined with prior information gathered from the herd Schaeffer and Jamrozik (1996) suggested a multiple-trait procedure that incorporated information about average lactation curves and covariances between test day yields for milk, fat and protein Although these sophisticated methods are able to give predictions with reasonable accuracy, it is at the expense of great computational demand and mathematical effort (Macciotta et al , 2002)

Thus, the empirical regression approach to modelling lactation curves is more effective for prediction purposes (Perochon et al , 1996) Often, empirical regression methods are biologically interpretable and are considered more appropriate for use by bio-economists who need to constantly update and re-create the parameters for different scenarios The simplicity of use and the ability of empirical regression models to be biologically interpretable are the reasons why this study focuses on investigating the empirical regression methods of modelling lactation curves in the Irish context

## 2.3 Modelling Liveweight

Three approaches have been used to model the evolution of liveweight of cows modelling liveweight using body measurements (Gravir, 1967, Heinrichs et al , 1992, Wicks, 2001, Madalena et al , 2003), 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 , 2005) and modelling liveweight over a lactation (Wood et al , 1980, Korver et al , 1985, Lopez-Villalobos et al , 2001)

### 2 3 1 Estimating Liveweight Using Body Measurements

The motivation for research into the use of body measurements to predict liveweight is that most farms do not have weighing machines capable of measuring dairy cows' liveweight As liveweight depends largely on the content of rumen, intestinal canal and bladder, body measurements, and especially heart girth, are frequently used in order to estimate liveweight in cattle and are less likely to be affected by extraneous environmental factors (Gravir, 1967, Madalena et al , 2003) Gravir (1967) and Heinrichs et al (1992) found heart girth to be the best single estimator of liveweight and using one or more linearly combined measurements in addition to heart girth was found not to improve the estimate appreciably over using heart girth alone Wicks (2001) examined the models of Ayala et al (1992) and Devir et al (1995) and proposed a model, which included body condition score, body size, parity and stage of gestation, this model takes the following form

$$LW = -751 + 58\,7BCS + 7\,92Ht + 18\,3SG + 42\,7P \qquad (2\,3\,1)$$

where $LW$ is the liveweight, measured in kilograms, $BCS$ is body condition score, $Ht$ is the height at withers, $SG$ is the stage of gestation ($0$ = non-pregnant, $1 = 1$ to 90 days pregnant, $2 = 91$ to 180 days pregnant, $3 \geq 181$ days of gestation) and $P$ is the lactation number or parity ($1, 2$ or $\geq 3$) The model of Wicks (2001) and other similar functions consisting of body measurements assume that the body measurements are taken accurately but inevitably errors are frequently made when recording these measurements

### 2 3 2 Liveweight from Birth to Maturity

Modelling liveweight from birth to maturity was initially examined by Brown et al (1976) They investigated the well-known growth models (See Table 2 1) of Gompertz (Winsor, 1932), Brody (1945), Von Bertalanffy (1957) and Richards (1959) However, it was found that these consistently overestimated weight at early ages

Table 2 1 Equations for five growth curves

| Model | Equation[a] |
|---|---|
| Von Bertalanffy | $y_t = A(1 - Be^{-Kt})^3$ |
| Brody | $y_t = A(1 - Be^{-Kt})$ |
| Gompertz | $y_t = y_0 eL(1 - e^{-\alpha t})/\alpha$ |
| Richards | $y_t = A(1 - Be^{-Kt})^M$ |

[a] $y_t$ =weight at age $t$, $A, B, K, L, M$ and $\alpha$ are fitted parameters

Brown et al (1976) found that when fitting weight-age data where the goodness-of-fit, especially prior to ten months of age, is critical, a four parameter model, known as the Richards model (Richards, 1959) was most satisfactory This corresponds to the findings of Perotto et al (1992) Other authors, such as Moore (1985), proposed alternative methods to predict the body weight of a cow from embryo to adult using a model of the following form

$$W = A(1 + e^{-p_n \log_e(t-3.5)/A^{0.27}})^{-1/0.27} \tag{2.3.2}$$

where $W$ is the weight of the cow at age $t$, $p_n$ is an $n^{th}$ order polynomial and $A$ is a fitted parameter. There is no general consensus on the most universally suitable growth curve (Berry et al., 2005) but the function selection should depend upon the nature of the study and the intended application of the results.

### 2.3.3 Modelling Liveweight over a Lactation

The fore-mentioned models that describe liveweight are useful for describing the individual liveweight curve for more than one lactation. However, it is also necessary to model the liveweight changes within a lactation for the overall management of dairy herds and the formulation of diets. A combination of growth, change in alimentary tract fill, pregnancy and alternate deposition and catabolism of body tissue reserves have an effect on liveweight (Korver et al., 1985) which causes the curve of liveweight to fall rapidly after calving and then rise slowly until the next calving.

Wood et al. (1980) described liveweight changes of British dairy cows from several breeds using the gamma function (outlined in Section 2.2.1), but the analysis was restricted to the first 20 weeks after calving. Various researchers (Korver et al., 1985, Berglund and Danell, 1987, Lopez-Villalobos et al., 2001) have also used this model to describe the evolution of the liveweight of cows throughout a lactation. Korver et al. (1985) constructed a function, from the incomplete gamma function, incorporating liveweight level (scale) together with variables representing pregnancy status, the maximum decrease of liveweight during the lactation and the time during lactation at which minimum liveweight occurred as follows

$$y_{it_lt_p} = p_1 + p_2^3(t_p - 50)^3 + p_3t_l/p_4e^{(1-t_l/p_4)} + \epsilon_{it_lt_p} \quad (2\ 3\ 3)$$

where $y_{it_lt_p}$ is the liveweight of cow $i$ at $t_l$ days since calving and $t_p$ days pregnant, $p_1$ is the level of liveweight, $p_2$ is the pregnancy parameter, $p_3$ is the maximum decrease of liveweight during the lactation, $p_4$ is the time during the lactation with the minimum liveweight, $t_l$ is the number of days since calving, $t_p$ is the number of days pregnant ($t_p - 50 > 0$) and $\epsilon_{it_lt_p}$ is the error term Korver et al (1985) found it difficult to estimate the pregnancy parameter based on the measurements for the first 40 weeks of lactation, this being a major criticism of this work Berglund and Danell (1987) and Lopez-Villalobos et al (2001) used Wood's function to predict liveweight change, as this model had been used in earlier studies While Berglund and Danell (1987) focused their attention on the first period of lactation only, Lopez-Villalobos et al (2001) compared the lactation curves of milk traits, liveweight and body condition score for two genetic strains of cows, namely heavy and light Holstein-Friesians

### 2 3 4 Liveweight Literature Summary

The three approaches discussed to model the liveweight of cows each have their own qualities While the use of body measurements and growth curves present an overall picture of liveweight from embryo to maturity, a model which is sensitive to the rapid loss of liveweight after calving and then the gradual gain in liveweight until the start of the following calving is also required Accurate measures of liveweight during this period can be very beneficial when making management and nutritional decisions at herd level and for individual cows (Forbes, 1983, Walter et al , 1984) Therefore it is necessary to derive an equation in the Irish context that will model the liveweight change of dairy cows over a lactation with minimum error This model would need to take into account certain fundamental principles, such as age and lactation week Empirical regression methods such as those outlined by Wood et al (1980) and Korver et al (1985) formed the basis for investigating this in more detail

## 2.4 Chapter Summary

In this chapter an overview of some important research, which forms the foundation of the work presented in the rest of this thesis, was presented Several observations regarding the limitations of previous work have been made In order to model the lactation curves of milk yield, fat content, protein content and liveweight, the factors which affect these measures, the modelling techniques and the statistical analysis involved need to be addressed

From the review of the literature, in this chapter, it has been decided that empirical algebraic models will be examined when proposing modelling techniques to represent the shape of milk yield, fat content, protein content and liveweight curves As this study limits itself to examining these traits over a lactation period an empirical regression model would most likely yield the most useful results These models are often biologically interpretable and are easy to apply which is of great benefit to scientists and economists They are simple models and the literature suggests that they are generally more effective than more sophisticated models in predicting lactation curves The factors which affect the milk yield, fat content, protein content and liveweight of cows will be investigated in the next chapter before the modelling is performed

# Chapter 3

# General Analysis of Data

## 3 1 Introduction

This chapter describes the data used throughout this study and examines, for the first time in recent years, the factors which affect the milk production of Irish dairy cows It will analyse these factors and demonstrate the affects they have on milk yield, the concentration of its economically important constituents as well as the evolution of the liveweight of the animal throughout its lactation

A considerable amount of work has been done internationally in analysing the factors which affect milk production, but in the Irish context the issue has not been examined in detail for some considerable time Cunningham (1972) investigated the factors which effect total milk and fat yields of Irish dairy cows Killen and Keane (1978) again analysed the factors which affect total milk, fat and protein yields and those that affect the average fat and protein content in milk The factors affecting milk yield and its composition have not been examined since then and this topic therefore requires revisiting, as the Irish dairy industry has grown and has advanced quickly in recent times The factors affecting the liveweight of an animal over a lactation have not been examined in Ireland to date and this chapter addresses some of the issues involved

## 3 2 Milk Yield, Fat Content and Protein Content Data

The data available for this study came from commercial and experimental dairy herds which included both autumn and spring calving cows 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, collecting detailed data on stock, farm inputs, production, and reproduction information on a monthly basis The lactation number of each cow was obtained from the Irish Dairy Records Cooperative (IDRC) files and calving date records were captured through the DairyMIS system The calving date and drying date (also obtained from the IDRC files) were used to validate lactation number and test-day records for a given lactation The recordings consisted of cow identification number, lactation number, breed, year, week end date, week number, calving day of year, lactation week, milk yield (kg), fat yield (kg), protein yield (kg), lactose yield (kg), fat content, protein content, lactose content, liveweight (kg), condition score, breed, genotype and feed system Not all of these recordings were available for all animals but there was a sufficient number to enable a detailed analysis to be carried out

### 3 2 1 Datasets

Two sets of data were made available for this study, Dataset 1 consisted of 1,729 weekly records over the period 1995 - 2002 from 905 individual cows on experimental herds and Dataset 2 comprised 14,198 records[1] with monthly test-day yields recorded during 1999 and 2000 from 79 commercial spring-calving dairy herds (See Table 3 1) There were 15,606 records for spring/summer calving cows (SSC) and the remainder were for autumn/winter calving cows (AWC), defined as calving from July to December The autumn/winter calving cows were all in experimental herds The seasonal calving pattern of the animals in this dataset are shown in Table 3 2

[1]A record consists of approximately 44 recordings (test-day values) for Dataset 1, and approximately eleven recordings for Dataset 2

Table 3 1 Number of anımals per year

| Year | No of anımals | |
|---|---|---|
| | Experımental Herds[a] | Commercıal Herds[b] |
| 1995 | 161 | |
| 1996 | 271 | |
| 1997 | 232 | |
| 1998 | 207 | |
| 1999 | 297 | 7529 |
| 2000 | 344 | 6669 |
| 2001 | 214 | |
| 2002 | 3 | |

[a] Dataset 1
[b] Dataset 2

Table 3 2 Number of anımals per calvıng month ın each dataset

| Calvıng Month | No of Anımals | |
|---|---|---|
| | Experımental Herds | Commercıal Herds [a] |
| January | 157 | 1722 |
| February | 755 | 6012 |
| March | 378 | 3524 |
| Aprıl | 124 | 1719 |
| May | 25 | 444 |
| June | 2 | |
| July | 2 | |
| August | 2 | |
| September | 168 | |
| October | 57 | |
| November | 35 | |
| December | 24 | |

[a] 777 anımals ın commercıal herds for whıch calvıng data was not avaılable

The data consists of cows in lactations one through to sixteen However, all records for lactation number greater than two were grouped together (See Table 3 3) because it has been found that there is no significant difference between the behaviour of cows that are in their third lactation and those in later lactations (Cunningham, 1972, Killen and Keane, 1978, Vollebregt and Vollebregt, 1998, Lidauer et al , 2000, Dechow et al , 2004) The average yield per lactation was calculated only for Dataset 1 as shown in Table 3 4, because there were only part-lactation records available for Dataset 2 A cow in its first lactation produces, a total of, on average, 5,438 kg of milk while cows in third or higher lactation produce approximately 6,454 kg These would be reasonable approximations for commercial herds also The majority of cows in the study were Holstein-Friesian with the exception of 52 Normande cows and 55 Montebeliardes The average level for concentrate supplementation was 500 kg, with a range from 300 kg to 700 kg for the SSC cows, and approximately 1,500 kg for the AWC cows

These data were then used to examine the factors which affect total milk yield, fat content, protein content and liveweight As the liveweight data was often collected at a different time to the production data, liveweight was examined separately

Table 3 3 Number of animals per lactation

| **Lactation Number** | **No of animals** | |
|---|---|---|
| | **Experimental Herds** | **Commercial Herds** |
| 1 | 585 | 3508 |
| 2 | 484 | 3238 |
| 3+ | 660 | 7452 |

Table 3 4 Average milk yield per lactation(Experimental Herds)

| **Lactation No** | **Average Yield** | **Standard Deviation** |
|---|---|---|
| 1 | 5437 77 | 1024 75 |
| 2 | 6324 58 | 1276 03 |
| 3+ | 6453 77 | 957 66 |

### 3 2 2 Analysis of Milk Yield, Fat Content and Protein Content Data

In the first instance an analysis was carried out on the factors affecting milk yield and its constituents The analysis of variance (ANOVA) technique was applied using a general linear model of the following form -

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \quad + \beta_n X_n + \epsilon \qquad (3\ 2\ 1)$$

where $Y$ is the vector of dependent variable observations, $X_n$'s are independent variables, $\beta_n$ are the parameters to be estimated and $\epsilon$ is the vector of errors (Draper and Smith, 1981) The values of the parameters $\beta_0, \beta_1, \quad \beta_n$ are estimated using the method of least squares To perform ANOVA, the assumptions underlying it must be examined namely 1) equality of sample sizes and 2) homogeneity of variances In this chapter the sample sizes are equal, but the assumption of equal variances is violated When the assumption of equal variances is violated but the sample sizes are equal ANOVA is still reasonably robust (LeBlanc, 2004) and so was performed in this study Initially in this study, $Y$ represents the total milk yield per lactation, the average fat content or the average protein content $X_n$ are the categorical variables, those repeat 'factors' which may influence the dependent variable, $Y$, which are lactation number, calving month and the interactions between these factors

Firstly, an ANOVA was performed on total milk yield (Dataset 1 only) and the factors investigated were lactation number and calving month It is clear from the analysis of variance table (Table 3 5) that lactation number, calving month and the interaction of these are significant in explaining the variation in total milk yield The Type III SS (sum of squares) is the sum of squares for a balanced test of each effect, adjusted for every other effect, thus, it tests how well the model as a whole (adjusted for the mean) accounts for the dependent variable's behaviour It is widely acknowledged that lactation number has a substantial affect on milk yield

Table 3 5 Analysis of variance of total milk yield

| Source | df | Type III SS | F Value | Pr > F |
|---|---|---|---|---|
| Lactation Number | 2 | 58537816 9 | 15 39 | < 001 |
| Calving Month | 11 | 406145543 2 | 19 41 | < 001 |
| Lactation Number x Calving Month | 16 [a] | 154118290 3 | 5 06 | < 001 |

[a] Not all combinations of lactation number x calving month were available

$R^2$=0 241

Killen and Keane (1978) accounted for this by observing that animals in later lactations have steeper lactation curves, which rise to a higher peak resulting in a higher total production However, after lactation number three, a cow matures and her performance stabilises thereafter (Vollebregt and Vollebregt, 1998) Calving month has a very obvious affect on lactation yield after accounting for the other variables Late spring calvers are renowned for having shorter lactations because they are taken indoors late in their lactation and would require expensive supplementary feed to remain in production (Killen and Keane, 1978) The interaction effect is also significant, though, the *F-value* is 5 06 which is quite low Lactation number, calvmg month and their interaction effect account for 24 per cent ($R^2$ = 0 241) of the variation in the total milk yield

When examining the average fat and protein content, Dataset 1 and Dataset 2 were amalgamated, as the recordings for the commercial herds are evenly spread throughout the year an average fat and protein content value could be calculated It is interesting to see that only calving month is significant (at a five per cent significance level) in explaining both the fat content (see Table 3 6) and protein content (see Table 3 7) of milk This is in agreement with the results obtained by Wood (1976) and Killen and Keane (1978), which found that the fat content of milk is not significantly affected by lactation number Calvmg month accounts for just 1 8 per cent ($R^2$ = 0 018) and 2 4 per cent ($R^2$ = 0 024) variation in fat and protein content, respectively The $R^2$ value was much lower for the fat and protein content than the $R^2$ value associated with total milk yield, but this may be partly

Table 3 6 Analysis of variance of fat content

| Source | df | Type III SS | F Value | Pr > F |
|---|---|---|---|---|
| Lactation Number | 2 | 0 00005416 | 1 17 | 0 3106 |
| Calving Month | 11 | 0 00334857 | 13 14 | < 0001 |
| Lactation Number x Calving Month | 16 | 0 00042349 | 1 14 | 0 3076 |

$R^2$=0 018

Table 3 7 Analysis of variance of protein content

| Source | df | Type III SS | F Value | Pr > F |
|---|---|---|---|---|
| Lactation Number | 2 | 0 00003902 | 2 09 | 0 124 |
| Calving Month | 11 | 0 00208374 | 20 26 | < 0001 |
| Lactation Number x Calving Month | 16 | 0 00023621 | 1 58 | 0 655 |

$R^2$=0 024

explained by the size of the dataset The dataset for estimating the factors which effect milk yield consisted of 1,729 records (as total milk yield was only recorded for experimental herd) whereas that used for estimating the factors affecting fat and protein content consisted of 15,927 records

The breed and the feeding regime were also examined as factors which might affect milk yield and its constituents However, the data for this analysis is limited to a subset of Dataset 1 and specifically those that are SSC There were three different breeds included in this dataset namely Holstein-Friesian, Montebeliarde and Normande The subset of records chosen for this analysis were all associated with one specific experimental farm and all animals were subject to the same feeding regime There were 122 Holstein-Friesian cows, 52 Montebeliarde cows and 55 Normande cows in this subset Breed is a nominal variable and thus the approach to analysing its effect must be different to that used for factors such as calving month and lactation number The analysis initially focused on examining the effects of lactation number, calving month and the interaction between these factors on the subset of the data in which breed was specified, so that the $R^2$ values (for including and excluding breed) could be compared The $R^2$ values are much higher, in general, in this section as the dataset is much smaller with only 548 records From Table 3 8 it can be seen that lactation number, calving month and breed are significant in explaining total milk yield From the difference in $R^2$ values, between including and excluding breed, it shows that breed accounts for 21 per cent of the total variation in milk yield Similarly, breed accounts for eight per cent of the variation in the fat content of milk and 23 per cent of the variation in the protein content of milk produced

There were three different feeding regimes carried out on 214, 168 and 126 lactations of the same breed on a particular farm belonging to Moorepark Research Centre The feeding systems are outlined in Table 3 9 and it can be seen that these feeding systems held the nitrogen fertilisation rate constant while the stocking rate and concentrate input varied Given the nature of these feeding regimes, feed was examined, in a similar manner to breed, as a factor which affects milk yield and its constituents Table 3 10 shows that feed is also significant in explaining total milk yield Again by examining the difference in $R^2$ values, when feed is included or excluded, it shows that feed accounts for ten per cent of the total variation in milk yield Similarly, feed accounts for approximately four per cent of the variation in the fat content of milk and approximately five per cent of the variation in the protein content of milk

The estimates of the parameters were then found, for each categorical variable which was deemed significant, using the following model -

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \qquad (3\ 2\ 2)$$

where $Y$ is either the total milk yield or the average of one of its constituents, $x_1, x_2, x_3,$ etc are the categorical variables representing the factors being assessed

Table 3 8 Comparison of analysis of variance of milk yield excluding and including breed

| Source | Excluding Breed | | | | Including Breed | | | |
|---|---|---|---|---|---|---|---|---|
| | df | Type III SS | F Value | Pr > F | df | Type III SS | F Value | Pr > F |
| Lactation Number | 2 | 23587521 58 | 15 30 | < 001 | 2 | 34655792 34 | 29 30 | < 001 |
| Calving Month | 5 | 35445517 43 | 9 20 | < 001 | 5 | 33885138 09 | 11 46 | < 001 |
| Breed | | | | | 2 | 29151938 67 | 24 65 | < 001 |
| Lactation Number x Calving Month | 9 | 6718580 42 | 0 97 | 0 4653 | 8 | 5322894 86 | 1 13 | 0 3444 |
| Lactation Number x Breed | | | | | 4 | 2033956 67 | 0 86 | 0 4879 |
| Calving Month x Breed | | | | | 8 | 2531030 57 | 0 54 | 0 8303 |
| Lactation Number x Calving Month x Breed | | | | | 13 | 8787072 38 | 1 14 | 0 3198 |
| | | | | $R^2$=0 25 | | | | $R^2$=0 46 |

Table 3 9 Feeding systems

| **Feeding System** | **No of Animals** | **Stocking Rate** | **Nitrogen Fertilization Rate** | **Concentrate Input** |
|---|---|---|---|---|
| 1 | 221 | High (3 0 cows $ha^{-1}$) | High (380 kg N $ha^{-1}$) | 500 kg per cow over total lactation |
| 2 | 168 | High (3 0 cows $ha^{-1}$) | High (380 kg N $ha^{-1}$) | 1000 kg per cow over total lactation |
| 3 | 126 | Grazed to a higher post-grazing sward height reducing occupancy time per paddock | High (380 kg N $ha^{-1}$) | 500 kg per cow over total lactation |

Table 3 10  Comparison of analysis of variance of milk yield excluding and including feed

| Source | Excluding feed | | | | Including feed | | | |
|---|---|---|---|---|---|---|---|---|
| | **df** | **Type III SS** | **F Value** | **Pr > F** | **df** | **Type III SS** | **F Value** | **Pr > F** |
| Lactation Number | 2 | 54956390 06 | 25 15 | < 001 | 2 | 57277158 48 | 29 33 | < 001 |
| Calving Month | 4 | 43052174 74 | 9 85 | < 001 | 4 | 38603055 62 | 9 88 | < 001 |
| Feed | | | | | 2 | 24525312 88 | 12 56 | < 001 |
| Lactation Number x Calving Month | 7 | 22297277 84 | 2 92 | 0 0054 | 7 | 24522832 85 | 3 59 | 0 0009 |
| Lactation Number x Feed | | | | | 4 | 5195478 94 | 1 33 | 0 2576 |
| Calving Month x Feed | | | | | 7 | 5839439 65 | 0 85 | 0 5427 |
| Lactation Number x Calving Month x Feed | | | | | 11 | 11478808 22 | 1 07 | 0 3848 |
| | | | | $R^2$=0 33 | | | | $R^2$=0 43 |

and the interactions between these factors Table 3 5 shows that both lactation number and calving month were significant factors in explaining variation in milk production Using regression analysis, it was found that lactation number had a positive effect on total milk yield, from lactations one to three the total milk yield increases by approximately 468 kg per lactation (assuming the relationship is linear) Calving month does not appear to have a linear effect on yield and therefore it will be examined in detail elsewhere in this study (Chapter 5) As breed and feed are nominal categorical variables, and as such they do not have a natural ordering, they cannot be used in a linear equation (Agresti, 1990) Therefore, milk yield can be approximated by the following equation -

$$\text{Total Milk Yield} = \underset{(60\ 62)}{5108\ 55} + \underset{(12\ 22)}{468\ 17}(\text{lactation number}) \qquad (3\ 2\ 3)$$

$$R^2 = 0\ 0974$$

where the values in brackets are $t - statistic$ values

As milk processors pay producers on the basis of milk composition and not solely on milk quantity, the composition of milk is what needs to be optimised for farmers to receive their maximum profits A producer is also interested in predicting the liveweight of an animal as accurate estimates of liveweight are beneficial when making management and nutritional decisions

## 3 3 General Analysis of Liveweight Data

The liveweight data available comprised records from both experimental and commercial herds The dataset consisted of 6,455 records (each with approximately eleven recordings) taken at monthly intervals from 66 commercial herds, 334 records (each with approximately 44 recordings) taken at weekly intervals from six experimental spring calving herds and 94 records (each with approximately 44 recordings) from two experimental autumn calving herds over the period 1995-2001 The seasonal pattern of calvings of animals in this set of data is shown in Table 3 11 and it should be noted that there are no records available for cows calvmg in June, July or August These data included year of production, lactation number, calving month, lactation week and liveweight In the same way as in the examination of the milk yield data, lactation number was categorised into lactation 1, lactation 2 and lactation 3 or greater and Table 3 12 shows that each lactation group was well represented Liveweight was recorded by an automatic weighing system (DairyMaster) on the experimental farms, this system consisted of a scale with load cells and thus there was no visual recording of the weight On the commercial farms all cows in the herd were recorded electronically, using a portable weighing scales and Winweigh software The scales were calibrated weekly against permanent scales in Moorepark Research Centre and were calibrated again with known weights on arrival at each farm In all cases, recordings were taken after milking so as to minimise variations due to varying weights of gut fill

Table 3 11 Number of animals per calvmg month in the liveweight dataset

| Calving Month | No of Animals |
|---|---|
| January | 991 |
| February | 4435 |
| March | 1181 |
| April | 152 |
| May | 24 |
| June | - |
| July | - |
| August | - |
| September | 46 |
| October | 37 |
| November | 11 |
| December | 6 |

Table 3 12 Number of anımals per lactatıon ın the lıveweıght dataset

| Lactatıon Number | No of Anımals |
|---|---|
| 1 | 2206 |
| 2 | 1980 |
| 3+ | 2697 |

### 3 3 1 Analysıs of Lıveweıght Data

The hveweıght data were analysed ın a sımılar way to the mılk yıeld and mılk constıtuent data ın Sectıon 3 2 2, the ANOVA technıque beıng agaın used to examıne the factors whıch mıght affect the lıveweıght of cows The Type III sums of squares were examıned for the effect on hveweıght of lactatıon number, calvıng month and the ınteractıon of these factors It ıs clear from Table 3 13 that all three of these effects were sıgnıficant and accounted for approxımately 40 per cent of the varıatıon ın the average lıveweıght of the anımals

The coefficıents of the varıables were then found through general lınear modellıng It was found that calvıng month ıs not hıghly sıgnıficant ın explaınıng the average lıveweıght wıth a $t - statıstıc$ value (ın brackets) of 1 81 as follows

$$\text{Average Liveweight} = \underset{(257\ 17)}{443\ 022} + \underset{(74\ 29)}{47\ 410}(\text{lactation number}) + \underset{(1\ 81)}{0\ 85}(\text{calving month}) \quad (3\ 3\ 1)$$

$R^2$=0 393

Table 3 13 Analysıs of varıance of hveweıght

| Source | df | Type III SS | F Value | Pr > F |
|---|---|---|---|---|
| Lactatıon Number | 2 | 189510 14 | 38 52 | < 001 |
| Calvıng Month | 8 | 138684 96 | 7 05 | < 001 |
| Lactatıon Number x Calvıng Month | 14 | 82044 86 | 2 38 | 0 0026 |

$R^2$=0 401

Thus, as lactation number increases from one to three the liveweight of a cow increases by approximately 47 kg and cows that are in lactations greater than three will have a relatively constant average liveweight from lactation to lactation

## 3 4 Summary

This chapter aimed at analysing the factors affecting the milk yield, the milk constituents and the liveweight of Irish dairy cows in order to identify which categorical variables were significant In general, it was found that calving month, lactation number, breed and feed are significant factors which influence total milk yield while the fat and protein contents of milk are affected by calving month, breed and feed Thus, the results of fitting the general linear model confirm the well-documented effects of certain categorical variables in explaining variation in milk yield, fat content, protein content and liveweight Factors which need to be taken into account can therefore be identified when making predictions of milk yield, its constituents and liveweight

# CHAPTER 4

# DETECTING ABNORMAL DATA

## 4 1 Milk Recording in Ireland

At present in Ireland 391,975 cows out of a total dairy population of 1 173 million cows are milk recorded (ICBF, 2003) This represents 33 4 per cent of cows but only 20 6 per cent of dairy herds Currently, all milk recording in Ireland is conducted manually by seven milk-recording organisations It is common practice for sick cows not to be recorded but their test values are either predicted by the farmer or by the "recorder" based on previous test values, or else their test day values are declared "missing" In 2004, the Irish Cattle Breeders Federation (ICBF) together with Dairygold Co-operative conducted a pilot scheme which involved the introduction of an electronic do-it-yourself (DIY) milk recording system (ICBF, 2004) Some 140 herds participated in this pilot study and the results showed that there were lower overhead costs involved, less steps from collecting data to database and the system is easier to manage than in manual milk recording schemes (ICBF, 2004) As a result the introduction of this system may encourage recruitment of farmers to milk recording schemes The intention is to extend this system of milk recording nationwide in the near future The data handler was created by the New Zealand company, Tru-Test (ICBF, 2004), and it links the meter/milk jar readings with the cow's ear-tag number and its bar code, thus there is no manual sample taking and sick cows are recorded like the other cows However, the fact that a cow is sick represents only one reason for an abnormal recording

Abnormal recordings are defined as recordings that deviate significantly from the cow's other recordings Abnormally low or high recordings could be a result of injury or recording errors, such as meter malfunctions or improper sampling, as well as from data entry errors or incorrect identification (Guthrie, 1994, Slater and Webster, 2001b,a) Currently, there is no scientific method in place to detect abnormal recordings in Irish milk-recorded data, other than manual observations by the dairy farmer or the milk "recorder" (ICBF, 2005) Thus, an appropriate scientific method for detecting abnormal recordings is required, particularly if electronic milk recording devices are to be introduced in the future

The International Committee for Animal Recording stated in their revised recording guidelines (ICAR, 2002) that true daily test values collected from animals labelled by the farmer as sick, injured, under treatment or in heat must be used in the computation of the lactation record unless the test value is less than 50 per cent of the previous test value or less than 60 per cent of the predicted test value, if such is the case, these test values may be considered as missing Wiggans et al (2003) proposed a method for detecting and adjusting abnormal test day yields at the Animal Improvement Programs Laboratory in Maryland, USA This chapter will examine the method of Wiggans et al (2003) in the Irish context, it will investigate, for the first time, whether this method is effective in detecting abnormal recordings of fat and protein content and it will examine whether there are more abnormal recordings in commercial data than in experimental data

## 4 2 Model Development

The data used in this chapter consisted of a total of 15,927 lactations from two sets of data Dataset 1 comprised 1,888 lactations of weekly test day yields, from six experimental herds attached to Teagasc Dataset 2 comprised 14,198 lactations of monthly test day recordings from 79 commercial spring-calving dairy herds Records with fewer than five recordings were deleted from Dataset 2 and lactations of less than 25 weeks duration were removed from Dataset 1 After editing, Datasets 1 and 2 consisted of 1,727 and 13,229 lactations respectively and they were amalgamated for the analysis

### 4 2 1 Definition of Upper and Lower Limits

The cut-off points which determine whether or not a recording is described as abnormal need to be established before the abnormal recordings are detected A lower limit of 60 per cent of the predicted test day value was chosen, complying with the guidelines of the International Committee for Animal Recording (ICAR, 2002) An upper limit of 150 per cent of the expected value was chosen to capture the most extreme values As cows are more likely to produce an abnormally low yield rather than an exceptionally high yield (Wiggans et al , 2003) the limits are designed to take this into account

### 4 2 2 Estimation of Slope Parameters

Before detecting the abnormal recordings, the parameters for estimating the slope of the lactation curves needed to be calculated These were calculated separately for milk yield, fat content and protein content Second and subsequent test day slope values were calculated from the preceding test day using the following expression

$$(p_i - p_{i-1})/(TD_i - TD_{i-1}) = b_0 + b_1 TD_{i-1} + b_2 TD_{i-1}^2 + b_3 p_{i-1} + b_4 TD_{i-1} p_{i-1} + \epsilon \tag{4 2 1}$$

where $p_i$ = milk yield, fat content or protein content on test day $i$ ($TD_i$) Wiggans et al (2003) calculated the parameters by lactation stage ($< 50$ $DIM$ (Days in Milk) and $\geq 50$ $DIM$) and lactation number, however, in this study the parameters were calculated by lactation week (lactation week was chosen to replace lactation stage so that the estimation of the slope parameters would be more accurate), lactation number and calving month (calving month was included as it was deemed appropriate by Cunningham (1972) and also on the basis of the findings in Chapter 3) The slope between the first and second test day was calculated by using the subsequent test day value instead of the preceding test day in the following way

$$(p_1 - p_2)/(TD_1 - TD_2) = b_0 + b_1 TD_2 + b_2 TD_2^2 + b_3 p_2 + b_4 TD_2 p_2 + \epsilon \qquad (4\ 2\ 2)$$

where $p_i$ = milk yield, fat content or protein content on test day $i$ ($TD_i$)

### 4 2 3 Method for Detection of Abnormal Values

The predicted test day (apart from the first test day) value was calculated, using the slope parameters, as follows

$$\hat{p_i} = p_{i-1} + \hat{b}(TD_i - TD_{i-1}) \qquad (4\ 2\ 3)$$

where $\hat{p_i}$ = predicted milk yield, fat content or protein content on test day $i$, $p_{i-1}$ = observed normal milk yield, fat content or protein content on the preceding normal test day ($TD_{i-1}$) and $\hat{b} = b_0 + b_1 TD_{i-1} + b_2 TD_{i-1}^2 + b_3 p_{i-1} + b_4(TD_{i-1})(p_{i-1})$ using $b_0$, $b_4$ as estimated above If there was no preceding normal recording, $TD_i$ was tested against the herd mean value, adjusted for days in milk ($DIM$), calculated using least square means The first $TD$ value was tested against the second $TD$ value, if the second was declared normal as follows

$$\hat{p_1} = p_2 + \hat{b}(TD_2 - TD_1) \qquad (4\ 2\ 4)$$

where $\hat{p_1}$ = predicted milk yield, fat content or protein content on the first test day, $p_2$ = observed normal milk yield, fat content or protein content on the second test day and $\hat{b} = b_0 + b_1 TD_2 + b_2 TD_2^2 + b_3 p_2 + b_4(TD_2)(p_2)$ using $b_0$, $b_4$ as estimated in section 4 2 2 If the second test day value was declared abnormal then the first recording was tested against the herd mean value adjusted for $DIM$ (Wiggans et al , 2003)

The simplicity of this method is demonstrated as follows, if, for example, there were seven recordings for a certain cow, $r_1, r_2, r_3, r_4, r_5, r_6$ and $r_7$, as follows

$$\boxed{r_1} \longrightarrow \boxed{r_2} \longrightarrow \boxed{r_3} \longrightarrow \boxed{r_4} \longrightarrow \boxed{r_5} \longrightarrow \boxed{r_6} \longrightarrow \boxed{r_7}$$

each of the recordings, $r_2$ through to $r_7$, are compared to the preceding recording which has been declared normal If there is no preceding normal recording the recording being tested is compared to the herd mean value adjusted for $DIM$ However, as the first recording, $r_1$, has no preceding normal recording, it is compared to the second recording $r_2$, if $r_2$ is normal If $r_2$ is abnormal then $r_1$ is also compared to the herd mean value adjusted for $DIM$

## 4 3 Implementation of Model

While implementing this method for detecting abnormal recordings, a problem was encountered when the time lapse between the last normal recording and the one that was being tested was greater than one month, this problem had not been addressed by Wiggans et al (2003) When this was the case, the values outputted for the predicted milk yield, fat content or protein content were incorrect (they were increasing exponentially with time) if the method of Wiggans et al (2003) was used without adjustment To eliminate this problem, it was decided that if a time gap between recordings of greater than one month was encountered, then the recording being examined was compared with the herd mean value adjusted for $DIM$

## 4 4 Performance of Method

Using this method, three per cent of milk yield recordings were detected as being abnormal, for fat and protein content, the corresponding level of abnormal recordings was five per cent and one per cent, respectively. Table 4.1 shows the percentage of abnormal recordings per lactation category for milk yield, fat content and protein content. It shows that there is a relatively constant percentage of abnormal recordings in each lactation number category. The distribution of abnormal recordings for milk yield, fat content and protein content are given in Figures 4.1, 4.2 and 4.3, respectively. The percentage of the total number of abnormal recordings was found to be highest during early lactation for milk yield (Figure 4.1) and protein content (Figure 4.3) and during late lactation for fat content (Figure 4.2).

As would be expected, there were more milk yield and fat content recordings declared abnormal than protein content recordings as the protein content of milk is generally homogeneous (Klopcic et al., 2003). The highest percentage of abnormal recordings occurred in the first week of lactation for milk yield (See Figure 4.1). This is most likely because of the influence of nutritional regimes, the interval from calving to first test and lactation (familiarity with parlour, stress of the new environment for first time calvers, etc.). However, it may also be due to the inability of the cow, immediately post calving, to consume sufficient energy to sustain lactation (Mackle et al., 1999, Buckley et al., 2003, Butler et al., 2003, McGuire et al., 2004). There was found to be a constant percentage of abnormal milk yield recordings from week four to week 28 (approximately 1.3 per cent of recordings being abnormal), with 52 per cent of those abnormal recordings being under the lower limit value. This is possibly due to cows being sick or lame. For those abnormal recordings that are over

Table 4.1 Percentage of abnormal recordings per lactation for milk yield, fat content and protein content

| | **Percentage of Abnormal Recordings per lactation category** | | |
|---|---|---|---|
| | **1** | **2** | **3+** |
| Milk Yield | 3.26 | 2.59 | 3.91 |
| Fat Content | 4.16 | 4.46 | 6.33 |
| Protein Content | 0.33 | 0.19 | 0.26 |

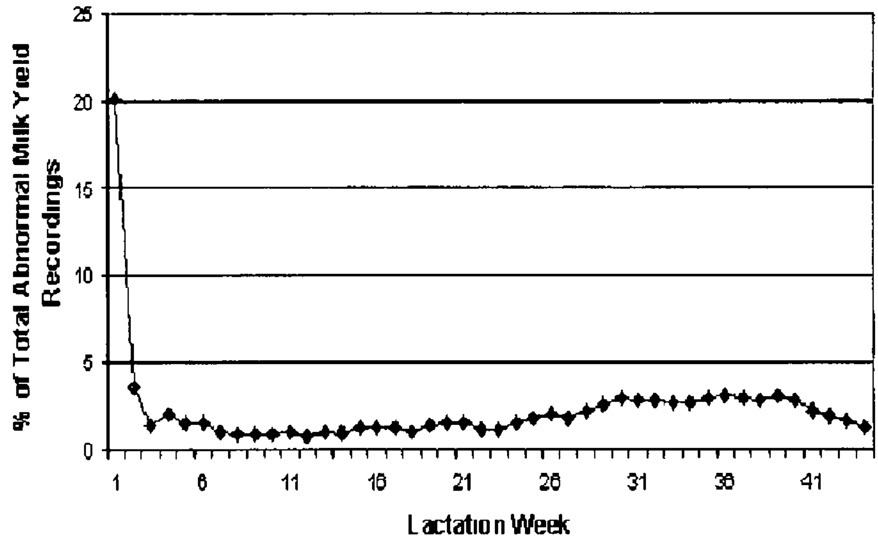


Figure 4 1 Percentage of total abnormal milk yield recordings per lactation week

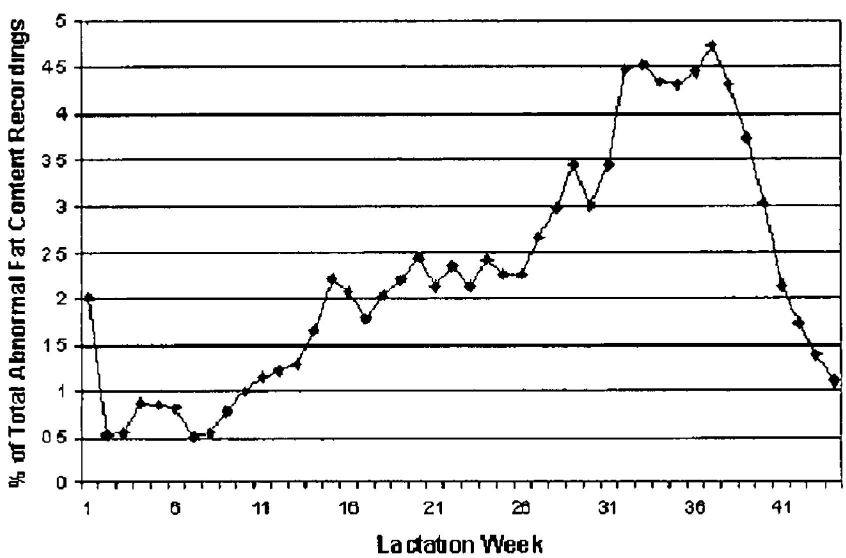


Figure 4 2 Percentage of total abnormal fat content recordings per lactation week

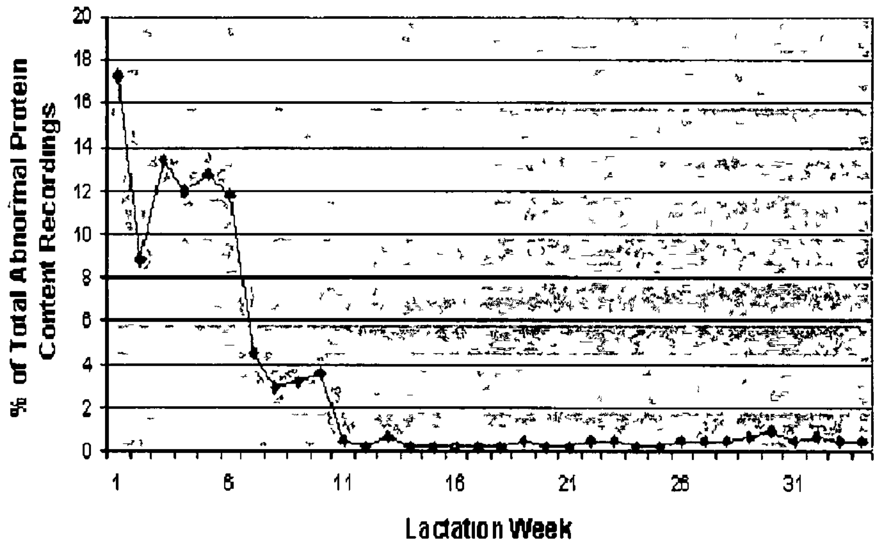


Figure 4 3 Percentage of total abnormal protein content recordings per lactation week

the upper limit value, the upper limit may in fact be eliminating some high yielding cows From lactation week 29 until end of lactation there is a slight increase in the number of abnormal recordings of which 57 per cent of the abnormal recordings are abnormally low Yields are most variable around this stage of lactation (Wiggans et al , 2003) again most likely due to nutritional regimes and management practices

In Figure 4 2, there is a constant slight increase (noting the scale of the graph) in the number of abnormal recordings until late in lactation, similar to the findings of Wiggans et al (2003) However, there is a severe peak in the number of abnormal recordings after lactation week 30, 97 per cent of these abnormal recordings are abnormally low This is probably influenced by factors such as the milk recorders not being sensitive enough to detect the rapid rise in fat content at this time in lactation or the milk not being agitated properly in the milk jars The drop in the percentage of abnormal recordings after week 36 is possibly due to there being fewer recordings in these weeks i e due to short lactations

The slight initial peak in Figure 4 3 is related to the composition of milk at this time Cows tend to have high milk protein in early lactation (Wood, 1976, Killen and Keane, 1978, Wilmink, 1987) and as a result a high number of abnormal recordings of which 69 per cent of the abnormal recordings in the first six weeks of lactation were declared abnormally high However, in total only 0 3 per cent of the protein recordings were declared abnormal

The number of abnormal recordings in the commercial herds and experimental herds were compared It was found that three per cent of the total experimental recordings and four per cent of the total commercial recordings were abnormal Similarly for fat content, one per cent of the experimental recordings and eight per cent of the commercial recordings were declared abnormal, while for protein content less than one per cent of the experimental and commercial recordings were declared abnormal It is interesting to note that there are significantly fewer abnormal recordings in the experimental herds than in the commercial herds This is possibly due to the fact that the experimental herds are predominately electronically recorded by a system that is approved by ICAR whereas the commercial herds are all manually recorded It will be even more interesting to investigate, in time, if the electronic DIY device will eliminate some of these abnormal recordings or if the number of abnormal recordings will actually increase

## 4 5 Chapter Conclusions

It is clear from the foregoing analysis that the method of Wiggans et al (2003) for detecting abnormal records is effective in the Irish context Some minor adjustments in relation to missing values particularly in commercial data that are recorded every four weeks were deemed necessary It was the first time that this method was used with fat and protein content data and again it proved satisfactory, with five per cent of fat content recordings and less than one per cent of protein content recordings being declared abnormal Considerably more abnormal recordings were found in the commercial data than in the experimental data but it will be interesting to investigate the impact, if any, the new electronic DIY recording device will have on these findings In conclusion, as there is no scientific method for detecting abnormal recordings currently in place in Ireland, it is desirable that a method for detecting abnormal recordings is established, especially as electronic devices are replacing manual recorders The method outlined in this chapter has been shown to be very effective in the Irish context and it abides by the guidelines outlined by ICAR (2002)

CHAPTER 5

# MODELLING MILK YIELD

## 5 1 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 from incomplete data Appropriate models provide useful information for breeding and management decisions at both industry and farm level 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 must accurately depict what is expected at farm level

Many authors have contributed to the evolution of research into lactation curves by using empirical regression models, test-day models, multiphasic models, Bayesian analysis and autoregressive procedures, as outlined in Chapter 2, this study will focus on empirical regression models Empirical regression models have been found to perform well statistically over a wide variety of datasets and they are often biologically interpretable They are easy to apply and therefore they are of great benefit to scientists and economists A summary of the empirical regression models investigated in this chapter is shown in Table 5 1 However, at present in Ireland, the SLAC

Table 5 1 Summary 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 | $ln(Y_n) = ln(a) + bln(n) - cn$ |
| 1971 | Dave | $Y_n = a + bn - cn^2$ |
| 1977 | Yadav et al | $Y_n = \frac{n}{a+bn+cn^2}$ |
| 1978 | Cobby and Le Du | $Y_n = a - bn - ae^{-cn}$ |
| 1979 | Madalena et al | $Y_n = a - bn$ |
| 1979 | Molina and Boschini | $Y_n = a - b\|n - c\|$ |
| 1982 | Singh and Gopal | $Y_n = a - bn + cln(n)$ |
| 1987 | Ali and Schaeffer | $Y_n = a + b\gamma + c\gamma^2 + d\omega + e\omega^2$ ** |
| 1987 | Wilmink | $Y_n = a + be^{-kn} + cn$ |
| 1995 | Guo and Swalve | $Y_n = a + b\sqrt{n} + clog(n)$ |

* $Y_n$ is the yield in lactation week $n$

** $\gamma = \frac{7n}{305}$, $\omega = ln\frac{305}{7n}$

(Standard Lactation Curve) method of Olori and Galesloot (1999) (as described in Section 2 2 6) is the preferred method for predicting milk yield but a single equation model would 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, improvements in cow production potential, and procedures to evaluate lactation curve models have resulted in a renewed interest in examining lactation curve models under Irish conditions Alternatives to the model of Wood (1967), 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 their respective studies The objectives of this chapter are to compare the goodness-of-fit of numerous empirical models including those of Wood (1967), Cobby and Le Du (1978), Ali and Schaeffer (1987), Wilmink (1987) and Guo and Swalve (1995) This chapter will also analyse the residuals arising from the fitting of each model in an attempt to find a well-fitting, robust, single equation model of weekly milk yields Thus, this work examines for each model its adherence to the assumptions of regression analysis in order to determine the reliability of the different models in estimating total milk yield tested under present day Irish production conditions If a model does not adhere to the assumptions of regression analysis it may not be the most reliable model to use as its robustness would be questionable An additional purpose of this analysis is to provide a seasonality production pattern table for use by bio-economists

The assumptions of a regression analysis that need to be examined to assess the appropriateness of a particular model are as follows

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 (homoskedasticity)

4 normality of the distribution of error terms

For a non-linear equation represented by the model $Y_u = f(\xi_u, \theta) + \epsilon_u$, where $Y_u$ is the dependent or response variable, $\xi_u$ is an independent or predictor variable $(u = 1, 2, \quad, n)$ and $\theta$ are the parameters, the true residuals, $\epsilon_1, \quad, \epsilon_u$, are assumed to be normally distributed ($N$) with zero mean and constant variance of $\sigma^2$

In Chapter 3 it was found that lactation number, calving month, herd, breed and feed had a significant effect on total milk yield per cow in Ireland and that the effect of these variables should therefore be taken into account Milk yield is also affected by environmental factors, such as weather and grass availability In particular, there is a stimulus to milk production due to 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 to examine this effect 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

## 5.2 Data used in this Study

Two sets of data have been made available for this study, as described in Chapter 3, Dataset 1 comprised lactation records from six research herds (Teagasc) recorded over the period 1995 to 2001 These data included 1,729 lactations from 872 individual cows, of which 1,408 lactations were from spring/summer calving (SSC) cows and the remainder were autumn/winter calving (AWC), defined as calving from July to December Dataset 2 comprised 14,198 lactations, with monthly test day yields recorded during 1999 and 2000 from 79 commercial dairy herds Within this dataset 4,336 cows had repeated lactations across the two years and all were spring/summer calving Records with fewer than five recordings were deleted from Dataset 2 and lactations of less than 25 weeks duration were removed from Dataset 1 Abnormal recordings were also detected, using the method outlined in Chapter 4 and removed from the datasets After edits, Datasets 1 and 2 consisted of 1,727 and 13,229 lactations, respectively

## 5 3 Investigation of Models

Models cited in the literature were investigated and the suitability of a model was evaluated on the basis of its goodness-of-fit to the data, its ability to satisfy the assumptions of regression analysis (namely autocorrelation, homoskedasticity, multicollinearity and normality of the distribution of error terms) and its ability to predict total milk yield These criteria were examined separately and then the overall best-fit model was chosen

### 5 3 1 Goodness-of-Fit

A preliminary examination was carried out on the models of Brody et al (1923), Brody et al (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 pooled data, using linear and nonlinear regression and then the effects of lactation number, calving month and herd were removed This was done by calculating the mean parameter estimates for each herd, and within each herd the mean parameter estimates for each calving month, and within each calving month finding the mean parameter estimates for each lactation number, these findings were used to give the mean parameter estimates for each model Those models which were found to be poor in fitting the Irish data (i e $R^2 < 0\ 60$) were eliminated from further consideration The $R^2$ value of a nonlinear equation was found by calculating

$$1 - \frac{SSE}{CSS} \tag{5 3 1}$$

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 et al , 1986) using the following formula -

$$MSPE = \frac{\sum_{t=1}^{n} e_t^2}{n} \tag{5 3 2}$$

where $e_t$ is the residual for observation $t$ and $n$ is the number of predicted values obtained This was calculated for each calving month and lactation number category within each herd before calculating the overall MSPE value for each model

This preliminary examination resulted in the models of Brody et al (1923), Brody et al (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 610), indicating a poor fit of these models to the data In fact, the model of Brody et al (1924) failed to converge using the Irish data The goodness-of-fit statistics of the expected curves for weekly milk yield, for the better-fitting models and for the model of Wood (1967), are presented in Table 5 2 It can be seen that the model of Ali and Schaeffer (1987) gave the best fit with a MSPE value of 501 79, while the model of Wood (1967), in weighted linear form, gave the poorest fit (MSPE value of 624 83)

The MSPE values found in this study for the linear and non-linear forms of the model of Wood (1967) reinforce the point made by Cobby and Le Du (1978) that non-linear regression would prove to be a more satisfactory method of fitting a nonlinear model After the preliminary examination, the models of Wood (1967) (non-linear form), Wilmink (1987), Ali and Schaeffer (1987) and Guo and Swalve (1995), were considered to have acceptable MSPE values and further analysis was performed on these models

### 5 3 2 Analysis of Residuals

Although many researchers have been involved in modelling milk yield using regression techniques, they did not test the validity of all the assumptions of regression analysis This study investigates the assumptions of regression analysis for the models cited in literature in the search to find the best empirical regression model of weekly milk yield throughout lactation in Irish dairy cows

Autocorrelation describes the situation where successive items in a series are correlated, so that their covariance is not zero and they are not independent This often occurs in time series data The Durbin-Watson statistic was calculated for

Table 5 2 Goodness-of-fit statistics of expected curves for weekly milk yield

| Model | MSPE value |
|---|---|
| Wood (Linear Form) | 583 99 |
| Wood (Weighted Linear Form) | 624 83 |
| Wood (Nonlinear Form) | 562 16 |
| Wilmink | 603 72 |
| Ali and Schaeffer | 501 79 |
| Ali-B | 520 93 |
| Guo and Swalve | 556 57 |

*Source* Quinn et al (2005a)

all of the models to test for the existence of autocorrelation between the residuals 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 study are those outlined by Mendoza (1999) (see Table 5 3) where $d_l$ is the lower critical bound and $d_u$ is the upper bound Initially, first order autocorrelation was examined and if this proved to be inconclusive then higher order autocorrelations were tested The Durbin-Watson statistic was found, in this study, to be between $d_u$ and $4 - d_u$ for first order autocorrelation, for all of the models investigated, indicating that autocorrelation was not present (Table 5 4)

If the residuals in the regression equation have a common variance the model is referred to as homoskedastic The test for violation of homoskedasticity is White's test (Sen and Srivastava, 1990) Heteroskedasticity is caused by the non-normality of one of the variables, an indirect relationship between variables, or can result from the effect of a data transformation Heteroskedasticity is not fatal to an analysis but the analysis is weakened when it is present White's test was calculated for each individual lactation, a mean value being computed after accounting for calving month, lactation number and herd effect 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 general test used to investigate whether or not the residuals of a regression analysis are normally distributed is the Kolmogorov-Smirnov test It was found

Table 5 3 Decision rules for autocorrelation

| **Null Hypothesis** | **Condition** | **Decision** |
|---|---|---|
| No positive correlation | $0 < d < d_l$ | Reject null (Autocorrelation) |
| No positive correlation | $d_l < d < d_u$ | No decision |
| No negative correlation | $4 - d_l < d < 4$ | Reject null (Autocorrelation) |
| No negative correlation | $4 - d_u < d < 4 - d_l$ | No decision |
| No autocorrelation | $d_u < d < d_l$ | Do not reject null |

Table 5 4 Comparison of models for milk yield

| Test | Wood | Wilmink | Ali and Schaeffer | Ali-B | Guo and 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 (Condition Index) | Weak (25 2) | Weak (15 3) | Strong (1075 4) | Moderate (55 9) | Moderate (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 |

*Source* Quinn et al (2005a)

using this test that the assumption of normality of the residuals was not violated as the $p-values$ for the Kolmogorov-Smirnov test statistic, $D$, varied from 0 10 to 0 11 across the models investigated (See Table 5 4) Additional measures which help to understand the distribution of the data are skewness (which identifies lack of symmetry in a given distribution) and kurtosis (which identifies weight in the extremes in a probability distribution), the kurtosis value of a normal distribution being zero Kurtosis values varied between 0 34 and 0 78, while skewness varied between -0 04 and -0 18 Thus, it was concluded from the examination of the Kolmogorov-Smirnov test statistic, kurtosis and skewness that there was no significant deviation from normality in the distribution of the residuals for the models of Wood (1967) (nonlinear form), Wilmink (1987) Ali and Schaeffer (1987) and Guo and Swalve (1995)

The situation where the explanatory variables are highly intercorrelated, thus making it difficult to distinguish the separate effects of the explanatory variables on the dependent variable, is known as multicollinearity A condition index was calculated to test for the presence of multicollinearity If the condition index values ranged from 30 to 100 it indicated that moderate to strong multicolhnearity existed (Belsley et al , 1980) When multicollinearity exists two problems may occur the computation of the parameter estimates may be slow and non-convergent, and the parameter estimates may have inflated variances (Belsley et al , 1980) To reduce the presence of multicollinearity at least one of the variables needs to be removed Examination of the multicollinearity diagnostics revealed that there was a strong presence of multicollinearity when applying the model of Ali and Schaeffer (1987), with a condition index value of 1075 4 (Table 5 4) In the model of Guo and Swalve (1995) there was moderate multicollinearity (condition index value of 49 39), whereas in the models of Wood (1976) and Wilmink (1987), multicollinearity was weak (condition index values of 25 20 and 15 32, respectively) As the model of Ali and Schaeffer (1987) has the best MSPE value but the most severe problem of multicolhnearity, this model was then examined with each of the variables removed ın turn (thus removıng a parameter each tıme the model was re-estımated) It was found that the condıtıon ındex could be reduced when parameter $b$, $c$, $d$ or $e$ was removed, but the greatest ımprovement occurred when parameter $b$ (assocıated wıth the $\gamma$ varıable) was removed The MSPE value for thıs new model (wıthout the $b$ parameter), denoted the Alı-B model, was 520 93, whıch ıs margınally hıgher than the MSPE value found when fittıng the orıgınal model of Alı and Schaeffer (1987)

As a result of these regressıon analyses, the models used to estımate the expected lactatıon curves for mılk yıeld are shown ın Table 5 5 The estımates of the $b$ parameter for the three forms of Wood's model were very sımılar 0 35, 0 27 and 0 32 for Wood's lınear, weıghted lınear and nonlınear forms, respectıvely The values were also very sımılar for the estımates of the parameter $c$, between the three estımatıon procedures (0 041, 0 035 and 0 039, respectıvely)

A comparıson of the parameter estımates for the model of Wood (1967) ın the study of Kıllen and Keane (1978) wıth those found ın thıs study gıve an ındıcatıon of changes whıch have occurred ın daıryıng ın the 27 years between the two studıes (See Fıgure 5 1) The mean values of the shape parameters have ıncreased from 0 331 to 0 353 for parameter $b$ and have decreased slıghtly (ın absolute terms) from -0 058 to -0 041 for $c$ In the 1978 study, peak yıeld was estımated to have occurred around week sıx of lactatıon, whereas ın thıs study ıt was observed around week eıght

Table 5 5 Expected curve models for mılk yıeld

| **Model** | **Equatıon estımated*** |
|---|---|
| Wood (Lınear Form) | $Y_n = exp(119\ 87 + 0\ 35 \ln n - 0\ 04n)$ |
| Wood (Weıghted Lınear Form) | $Y_n = exp(142\ 02 + 0\ 27 \ln n - 0\ 04n)$ |
| Wood (Nonlınear Form) | $Y_n = 143\ 13n^{0\ 32}e^{-0\ 04n}$ |
| Wılmınk | $Y_n = 262\ 37 - 102\ 46e^{-0\ 10n} - 4\ 53n$ |
| Alı and Schaeffer | $Y_n = 194\ 42 - 99\ 87\gamma - 16\ 20\gamma^2 + 19\ 35\omega - 7\ 92\omega^{2**}$ |
| Alı-B | $Y_n = 121\ 98 - 52\ 46\gamma^2 + 71\ 06\omega - 18\ 94\omega^{2**}$ |
| Guo and Swalve | $Y_n = 190\ 18 - 71\ 49\sqrt{n} + 95\ 20 \log n$ |

* $Y_n$ ıs the yıeld ın lactatıon week $n$

** $\gamma = \frac{7n}{305}, \omega = ln\frac{305}{7n}$

*Source* Quınn et al (2005a)

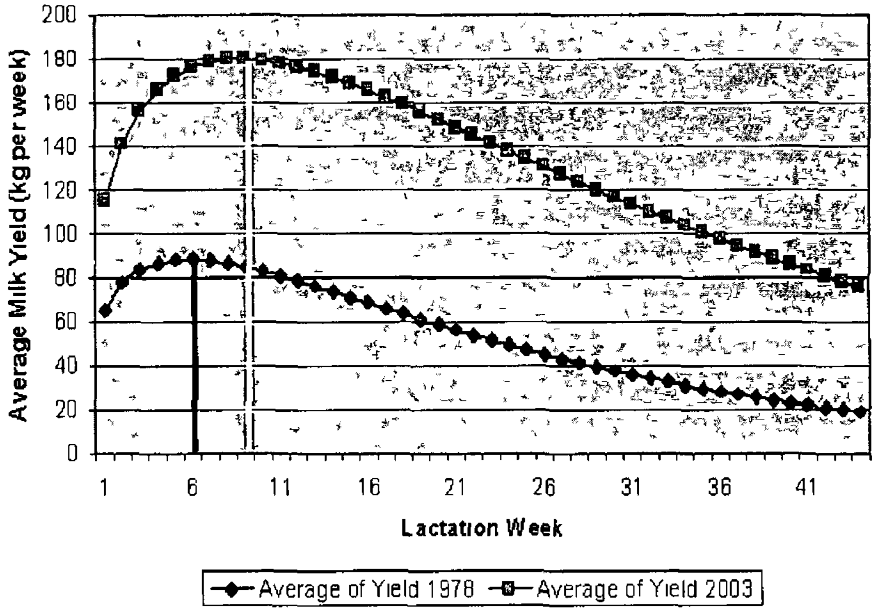


Figure 5 1 Comparison of lactation curves in 1978 and 2003, using the model of Wood (linear form), showing the peak week of milk yield

In comparison with the study of Killen and Keane (1978) the average milk yield per cow, as predicted by the model of Wood (1967), in linear form, has increased more than twofold from 2,364 kg to 5,448 kg between the two studies It must be acknowledged, however, that the two studies used quite different datasets, the dataset used by Killen and Keane (1978) came from experimental herds only, while the dataset in the present study, as well as being somewhat larger, includes data from commercial dairy herds

### 5 3 3 Estimation of Total Milk Yield

As the total milk yield is only known for the cows in Dataset 1, 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 in this dataset only This was achieved 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 of 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 average actual total milk yield for Dataset 1 was 5,702 kg The model of Ali and Schaeffer (1987) (See Table 5 6) gave an estimated total milk yield of 5,937 kg, overestimating the total by almost four per cent The Ali-B model estimated the total milk yield to be 5,795 kg, only overestimating the average by 1 6 per cent The models of Wood (1967), Wilmink (1987) and Guo and Swalve (1995) also overestimated total milk yield by 11 2 per cent, 1 9 per cent and 2 5 per cent respectively

## 5 4 Selection of Most Satisfactory Model

The main objective of this chapter has been to examine the suitability of a number of single equation models to explain the level of milk yield throughout a lactation The results show that the Ali-B model is the most consistent at adhering to the

Table 5 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 and Schaeffer | 5937 | 3 9 |
| Ali-B | 5795 | 1 6 |
| Guo and Swalve | 5849 | 2 5 |

*Source* Quinn et al (2005a)

regression assumptions and predicting individual weekly and total milk yields This model represents a considerable improvement of the linearised model of Wood which was used in the study of Killen and Keane (1978)

For the most promising models, the residuals 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 and an examination of kurtosis and skewness that there was no significant deviation from normality in the distributions of the residuals when analysing the models of Wood (1967) (nonlinear form), Wilmink (1987) Ah and Schaeffer (1987), Guo and Swalve (1995) and the model of Ali-B

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 the model of Ali and Schaeffer (1987), but when parameter $b$ was removed it was found that the resulting model (Ali-B) was the most satisfactory in that multicollinearity was no longer a major issue and it satisfied all of the remaining assumptions While there was still some correlation 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 to ensure adherence to regression assumptions Ali and Schaeffer (1987) included the $b$ parameter and the variable associated with it ($\gamma$, where $\gamma = \frac{7n}{305}$) to give a better fit to the data However, they failed to note that multicollinearity was present at an unacceptable level The MSPE value for the Ali-B model was 520 93, which is slightly higher than the MSPE value of the original model of Ali and Schaeffer (1987) but the problem of multicollinearity among the independent variables inflating the standard errors has been reduced

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 less than two per cent of the actual milk yield However, the prediction of total yield using the model of Ali and Schaeffer (1987) deviated from actual total milk yield by almost four per cent

### 5 4 1 Examination of Parameter Estimates and the Effects of Lactation Number, Breed and Feed

The parameter estimates were calculated for the Ali-B model for each lactation group A one-way analysis of variance (ANOVA) was then performed to investigate if the parameter estimates for each lactation category were significantly different from each other One-way ANOVA avoids the Type I error inherent in performing multiple $t-tests$, however the assumption of homogeneity of variances underlies ANOVA and therefore needs to be examined prior to performing this technique The Levene statistic (a measure of homogeneity of variance) was calculated (See Table 5 7), and it was found that there was no evidence to suggest that the variances were not equal ($p-values$ ranging from 0 284 to 0 454) for the parameter estimates Thus, ANOVA can be carried out to examine if the parameter estimates differ between lactation groups and if so, a post-hoc test assuming equal variances will show where these differences occur It can be seen from Table 5 8 that when using

Table 5 7 Test of homogeneity of variances for the parameter estimates of the Ali-B model for different lactation categories

| Parameter Estimate | Levene Statistic | df1* | df2** | p – value |
|---|---|---|---|---|
| a | 0 81 | 2 | 28 | 0 454 |
| c | 1 19 | 2 | 28 | 0 319 |
| d | 0 95 | 2 | 28 | 0 399 |
| e | 1 32 | 2 | 28 | 0 284 |

* $df1$ =degrees of freedom between groups
** $df2$ =degrees of freedom within groups

Table 5 8 One-way analysis of variance to compare the parameter estimates of the Ali-B model for each lactation group

| Parameter Estimate | Source | df | Sum of Squares | Mean Square | F Value | p – value |
|---|---|---|---|---|---|---|
| a | Between Groups | 2 | 17713 48 | 8856 74 | 0 232 | 0 795 |
| | Within Groups | 28 | 1069730 11 | 38204645 | | |
| | Total | 30 | 1087443 59 | | | |
| c | Between Groups | 2 | 38811 14 | 19405 57 | 0 169 | 0 845 |
| | Within Groups | 28 | 3212294 32 | 114724 80 | | |
| | Total | 30 | 3251105 46 | | | |
| d | Between Groups | 2 | 54621 94 | 27310 97 | 0 610 | 0 550 |
| | Within Groups | 28 | 1253733 83 | 44776 21 | | |
| | Total | 30 | 1308355 77 | | | |
| e | Between Groups | 2 | 6690 29 | 3345 14 | 0 859 | 0 435 |
| | Withm Groups | 28 | 109098 76 | 3896 38 | | |
| | Total | 30 | 115789 05 | | | |

ANOVA, it was found that there was no significant difference between the different lactation group categories and thus a post-hoc test was unnecessary

Although it was found in Chapter 3 that lactation number had a significant effect on total milk yield, this study shows that having adjusted for calving month and herd, the effect of lactation number is not significant among the parameter estimates for the Ali-B model This means that a single equation of the Ali-B model is used to describe the shape of a lactation curve across all lactations

As breed and feed were found in Chapter 3 to be significant factors affecting total milk yield, a similar analysis was conducted on the shape parameters in relation to these factors This analysis was confined to the subsets of data where breed and feed were specified Firstly, the parameter estimates having been adjusted for calving month, lactation number and herd were calculated for each breed, and a test for homogeneity of variances and ANOVA were performed The Levene statistic (Table 5 9) showed no evidence that the variances were not equal, at a five per cent significance level, for the estimated parameters $a$, $c$ and $d$, but this is not the case for parameter $e$ As the sample sizes of the groups are equal in this study a one-way ANOVA is reasonably robust even when the assumption of equal variances is violated (LeBlanc, 2004) It was found using ANOVA (Table 5 10) that there was a significant difference between parameter $c$ for the different breeds, although not highly significant ($p - value = 0\ 042$) The Dunnett's T3 post-hoc test assuming unequal variances (which keeps very tight Type I error control (Field, 2000)) was

Table 5 9 Test of homogeneity of variances for the parameter estimates of the Ali-B model for different breeds

| Parameter Estimate | Levene Statistic | df1 | df2 | p – value |
|---|---|---|---|---|
| a | 2 74 | 2 | 6 | 0 143 |
| c | 1 57 | 2 | 6 | 0 284 |
| d | 1 15 | 2 | 6 | 0 377 |
| e | 15 83 | 2 | 6 | 0 004 |

Table 5 10 One-way analysis of variance to compare the parameter estimates of the Ali-B model for each breed

| Parameter Estimate | Source | df | Sum of Squares | Mean Square | F Value | p − value |
|---|---|---|---|---|---|---|
| a | Between Groups | 2 | 1508 54 | 754 27 | 1 88 | 0 232 |
| | Within Groups | 6 | 2403 01 | 400 50 | | |
| | Total | 8 | 3911 55 | | | |
| c | Between Groups | 2 | 16119 76 | 8059 88 | 5 60 | 0 042 |
| | Within Groups | 6 | 8633 04 | 1438 84 | | |
| | Total | 8 | 24752 80 | | | |
| d | Between Groups | 2 | 5648 45 | 2824 23 | 3 53 | 0 097 |
| | Within Groups | 6 | 4798 75 | 799 79 | | |
| | Total | 8 | 10447 21 | | | |
| e | Between Groups | 2 | 532146 51 | 266073 26 | 1 03 | 0 421 |
| | Within Groups | 6 | 1549556 93 | 258259 49 | | |
| | Total | 8 | 2081703 44 | | | |

used to analyse these differences in more detail Table 5 11 shows that in fact the parameter estimates do not differ significantly between the breeds As the assumption of homogeneity of variances was violated, the results of Dunnett's T3 test are more reliable than those of ANOVA Similarly, the parameter estimates were calculated for each feeding regime outlined in Chapter 3 It was found that the variances of the parameter estimates could be assumed equal for each feeding regime using Levene's statistic and that there was no significant difference between the lactation categories for each parameter estimate (Table 5 12)

## 5.5 Application of Proposed Model

Overall, the results show that a modified version of the model of Ali and Schaeffer (1987) (Ali-B) best meets the criteria of predicting the weekly yields of individual cows The Ali-B model is a polynomial regression model of the following form

$$Y_n = a + c\gamma^2 + d\omega + e\omega^2 + \epsilon \tag{5 5 1}$$

where $Y_n$ is the yield in lactation week $n$, $\gamma = \frac{7n}{305}$, $\omega = \ln \frac{305}{7n}$, $\epsilon$ is the residual and $a$, $c$, $d$ and $e$ are the regression coefficients The resulting equation found for this particular model, as there is no significant difference between the parameter estimates for lactation number, breed or feed, is as follows

$$\begin{array}{lcccc} Y_n = & 121\ 83 & -52\ 46\gamma^2 & +71\ 06\omega & -18\ 94\omega^2 \\ & (19\ 04) & (26\ 71) & (1\ 56) & (2\ 05) \end{array} \tag{5 5 2}$$

where the values in brackets are the standard errors of the parameter estimates

The Ali-B model has been found to be the most robust single equation model in this study It is a single equation model and its implementation would be relatively simple when compared to the SLAC method that is currently being used for pre-

Table 5 11 Dunnett's T3 post-hoc test to find where the differences occur between lactations for each parameter estimate of the Ali-B model

| **Parameter Estimate** | **Breed (i)** | **Breed (j)** | **p – value** |
|---|---|---|---|
| a | Holstein-Friesian | Montebeliarde | 0 687 |
| | Holstein-Friesian | Normande | 0 126 |
| | Montebeliarde | Normande | 0 873 |
| c | Holstein-Friesian | Montebeliarde | 0 286 |
| | Holstein-Friesian | Normande | 0 159 |
| | Montebeliarde | Normande | 0 566 |
| d | Holstein-Friesian | Montebeliarde | 0 478 |
| | Holstein-Friesian | Normande | 0 186 |
| | Montebeliarde | Normande | 0 667 |
| e | Holstein-Friesian | Montebeliarde | 0 716 |
| | Holstein-Friesian | Normande | 0 712 |
| | Montebeliarde | Normande | 0 661 |

Table 5 12 One-way analysis of variance to compare the parameter estimates of the Ali-B model for each feeding regime

| Parameter Estimate | Source | df | Sum of Squares | Mean Square | F Value | p – value |
|---|---|---|---|---|---|---|
| a | Between Groups | 2 | 729 17 | 364 59 | 0 763 | 0 507 |
| | Within Groups | 6 | 2867 33 | 477 89 | | |
| | Total | 8 | 3596 51 | | | |
| c | Between Groups | 2 | 479 89 | 239 94 | 0 659 | 0 551 |
| | Within Groups | 6 | 2185 86 | 364 31 | | |
| | Total | 8 | 2665 75 | | | |
| d | Between Groups | 2 | 496 95 | 248 47 | 0 115 | 0 894 |
| | Within Groups | 6 | 13012 59 | 2168 77 | | |
| | Total | 8 | 13509 54 | | | |
| e | Between Groups | 2 | 22 23 | 11 12 | 0 111 | 0 896 |
| | Within Groups | 6 | 598 21 | 99 70 | | |
| | Total | 8 | 620 44 | | | |

diction purposes in Ireland However, while the parameters of the model of Wood (1967) are biologically interpretable, those of the Ali-B model are not, as it is a polynomial equation, but this does not negate its proficiency and usefulness It is also concluded that the Ali-B model is the best model for predicting weekly milk yield and it can also be easily used to create the seasonal production table for use by bio-economists

### 5 5 1 Derivation of Seasonal Effects and Production Pattern Table

Having determined the inherent shape of the lactation curve for an animal using the Ali-B model it was necessary to analyse them for environmental effects, regardless of stage of lactation The deviations resulting from comparing this 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 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 Table 5 13 shows the incremental change for these seasonal effects which should be added to the production effects given by the Ali-B model, independent of stage of lactation In May, for example, the Ali-B model underestimates the actual milk yield by almost nine per cent, due to environmental effects In December, on the other hand, the Ali-B model overestimates milk yield by over eight per cent It is evident that in the summer months there is a positive boost to milk production while in the winter months milk production is depressed below its expected level This trend 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

Table 5 13 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 |

*Source* Quinn et al (2005a)

To compute the percentage of the total lactation yield in each month throughout lactation, these seasonal effects were added to the best single equation model (Ali-B) An example of a production profile for a cow calving on the 15th day of each month is shown in Table 5 14 A cow calvmg in mid May, for example, will produce on average approximately five per cent of her total milk yield in May, 14 per cent of her total milk yield in June, 14 per cent in July and so on

Therefore, to calculate milk production for a particular animal for each production month, the Ali-B model is used to calculate the monthly potential yield, this is then adjusted by the seasonal effect in a particular month, as given in Table 5 13

## 5 6 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 model of Ah and Scha-

Table 5 14 An example of a productıon pattern profile for a cow calvıng on the 15th day of each month showıng the percentage of mılk yıeld suppled ın each productıon month

| | Calvıng 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 | 10 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 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 | 100 0 |

*Source* Quınn et al (2005a)

effer (1987), denoted Ali-B, 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 regression analysis procedures Although lactation number, breed and feed were found to be significant in explaining total milk yield (Chapter 3), the values of the estimates of the parameters which describe the shape of the Ali-B function, using one-way ANOVA, were found to be not significantly different from each other between these groups The Ali-B model is a relatively simple model to implement in practice when compared to the SLAC method which is currently in use in Ireland It has four parameters and can be fitted to any dataset using non-linear regression procedures 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, free of stage of lactation effects 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 most satisfactory model in the Irish context and it can be easily updated for different regional effects

# Chapter 6

# Modelling the Fat and Protein Content in Milk

## 6 1 Introduction

The concentration of fat or protein in the milk produced during a lactation can be represented by a curve, the shape of which normally mirrors a similar curve depicting milk yield (Pulina, 1990), the concentration of fat and protein in milk tends to decrease rapidly at the start of the lactation, and after falling to the minimum point, increases slowly until the lactation is completed (Wood, 1976) Typical fat and protein content curves are shown in Figure 6 1 The lowest point on the fat content curve lags approximately three weeks behind peak milk yield but in the case of protein concentration, it reaches its lowest point at approximately the same time as the peak in milk yield (Schutz et al , 1990, Stanton et al , 1992)

Algebraic models explaining production levels of the constituents in milk have many uses in farm management and in economic planning at farm level Farm models which can be used for component, system and management research require accurate data from a variety of sources The sub-models used within a farm system model need to be as accurate as possible in order for trustworthy farm models to be developed The Moorepark Dairy Systems Model (MDSM) (Shalloo et al ,

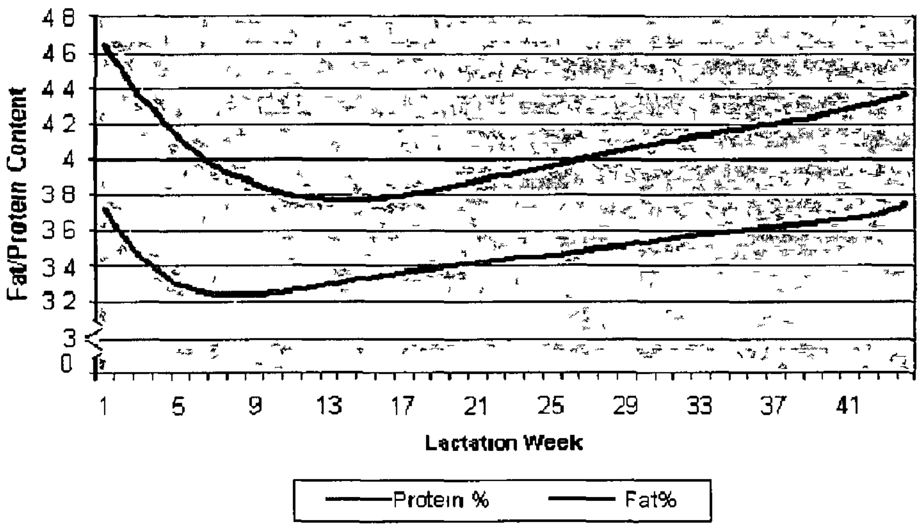


Figure 6 1 Typical fat and protein content curves

2004) is a farm simulation model which requires accurate descriptions of the fat and protein profiles of Irish cows At present in Ireland the preferred methodology to project partial lactation records, for milk yield and its constituents, is the SLAC (Standard Lactation Curve) method of Olori and Galesloot (1999) This is a method of interpolation consisting of a library of 2,160 equations for each constituent, from which the most appropriate equation is chosen The SLAC method is currently incorporated into the Moorepark Dairy Systems Model (MDSM), but the model is presently restricted to spring-calving cows The SLAC method projects weekly fat and protein yield, most of which is explained simply by the volume of milk produced It would be of great benefit to dairy scientists if a single equation model could explain as much of the variation in fat and protein content as the 2,160 equations proposed by Olori and Galesloot (1999) Recent research has also focused on using milk constituents to predict energy balance and fertility in high yielding cows (Heuer et al , 1999, De Vries and Veerkamp, 2000, Reksen et al , 2002, Buckley et al , 2003) where milk protein content or milk protein fat ratio are often used as indicators of energy balance (Fulkerson et al , 2001) The ability to accurately predict both the fat and protein content in milk is therefore also pertinent to fertility studies

This study investigates a number of empirical algebraic models and proposes a model that involves using fewer equations, compared with the SLAC method, to predict the fat and protein concentration of milk While some studies have analysed fat and protein yield (Wilmink, 1987, Stanton et al , 1992, Gonzalo et al , 1994, Schaeffer and Jamrozik, 1996, Jamrozik and Schaeffer, 1997, Olori and Galesloot, 1999, Garcia and Holmes, 2001, Vasconcelos et al , 2002) others have examined the fat and protein content (i e fat or protein percentage) of milk (Wood, 1967, Killen and Keane, 1978, Crosse et al , 1988, Stanton et al , 1992) The quality of milk is better explained by the fat and protein content than the fat and protein yield Thus, this chapter models the fat and protein content of milk for the first time in Ireland since the study of Crosse et al (1988)

As in the case of modelling milk yield, the work of Wood (1967) provides the starting point for many studies involving empirical algebraic equations for representing the fat and protein content curves

$$Y_n = an^{-b}e^{cn} \qquad (6\ 1\ 1)$$

where $Y_n$ is the fat or protein concentration of milk produced in week $n$ of lactation, $a$ is a scaling factor associated with the average fat or protein concentration, $b$ is related to pre-trough curvature and $c$ to post-trough curvature Several studies (Cobby and Le Du, 1978, Dhanoa, 1981, Rowlands et al , 1982, Ali and Schaeffer, 1987, Guo and Swalve, 1995) have found a poor fit to data on milk yield and its constituents when using this model, this poor fit may, in some instances, be due to environmental factors such as feed, weather and pregnancy status The model of Wood (1967) is, however, still considered by many to be a basic reference for research in the evolution of lactation performance of livestock (Varona et al , 1998)

Chapter 5 found that examining the residuals of regression analysis was beneficial in order to arrive at a well-fitting robust curve and therefore the residuals of the regression analysis are examined in this chapter when fitting models to fat and protein content Finally, the effect of environmental and seasonal factors, such as weather conditions, grass quality, etc free of stage of lactation effect, are also examined

## 6 2 Data

The data used in this study consisted of a total of 15,927 lactations from two sets of data as described in Chapter 3 Dataset 1 comprised 1,729 lactations of weekly test day yields from six experimental herds attached to Teagasc Dataset 2 comprised 14,198 lactations of monthly test day recordings from 79 commercial spring-calving dairy herds Records with fewer than five recordings were deleted from Dataset 2 and lactations of less than 25 weeks duration were removed from Dataset 1 After editing, Datasets 1 and 2 comprised 1,727 and 13,229 lactations respectively and they were amalgamated for the analysis Abnormal fat and protein content recordings were then removed using the method described in Chapter 4 After removing the abnormal recordings, the combined dataset consisted of 156,365 recordings from 14,952 lactations

## 6 3 Models and Statistical Analysis

Some of the models that were tested in Chapter 5 in relation to milk yield can be considered for analysing fat and protein content However the model of Ali and Schaeffer (1987) and the Ali-B model (Chapter 5) are not appropriate for modelling fat and protein content curves as they are based on polynomial expressions that keep their concave shape The models that were tested in this chapter were those of Wood (1967), Wilmink (1987) and Guo and Swalve (1995) The model of Wood (1967) was tested in three different forms, as discussed in Chapter 2 in nonlinear form, linear form (in which Wood's equation is linearised by taking the natural logarithm of each side of the equation), and weighted linear form (in which the logarithm of the fat or protein content is weighted proportionally to the square of the fat or protein content, as appropriate) When examining the model of Wilmink (1987), it was found that the value of the parameter $k$ was consistent (with a value of 0 10) over lactation number, calving month and herd effect for both fat and protein content It was found, after carrying out an initial analysis of variance (Chapter 3) similar to that of Cunningham (1972), that calvmg month, herd, breed and feed had a significant effect on the average fat content of milk while calvmg month, lactation number, herd, breed and feed had a significant effect on the protein content of milk and therefore the effect of these factors had to be considered when performing this regression analysis

### 6 3 1 Goodness-of-Fit

The Mean Square Prediction Error (MSPE) was used as a measure of goodness-of-fit, as outlined by Kvanli et al (1986) and the analysis of residuals was performed in the manner outlined in Chapter 5 (Section 5 3 2) The MSPE values are presented in Table 6 1, they range from 0 221 to 0 233 for the models when fitted to fat content data and they are 0 054 or 0 055 when fitted to protein content data It can be seen in Table 6 1 that there was no significant difference between the MSPE values for each model in fitting fat and protein content data

Table 6 1 Goodness-of-fit statistics of expected curves for fat and protein content

| Model | Fat Content MSPE value | Protein Content MSPE value |
|---|---|---|
| Wood (Linear Form) | 0 224 | 0 055 |
| Wood (Weighted Linear Form) | 0 233 | 0 054 |
| Wood (Nonlinear Form) | 0 222 | 0 054 |
| Wilmink | 0 221 | 0 055 |
| Guo and Swalve | 0 222 | 0 054 |

### 6 3 2 Analysis of Residuals

The Durbin-Watson statistic was calculated for each model to test for autocorrelation of error terms, while White's test was used to examine the homoskedasticity assumption A condition index was calculated to test for multicollinearity among explanatory variables (Belsley et al , 1980), while normality of the distribution of the error terms was examined using the Kolmogorov-Smirnov test together with measures of kurtosis and skewness

The results of the analysis of the residuals are shown in Tables 6 2 (for fat content) and 6 3 (for protein content) It can be seen that there was no first order autocorrelation present in any of the models analysed, for either fat or protein content, as 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)) Also White's Test (Sen and Srivastava, 1990) was non-significant ($p > 0\ 05$) for all models indicating that the homoskedasticity assumption was satisfied

Examination of the multicollinearity diagnostics when fitting the model of Guo and Swalve (1995) to the fat and protein content data revealed that moderate to strong multicolhnearity was present, indicated by condition index values of 89 6 and 87 0, respectively (A condition index value of between 30 and 100 indicates the presence of moderate to strong multicollinearity (Belsley et al , 1980)) For fitting the model of Wood (1967), in its three forms, there was moderate multicollinearity present, with condition index values of approximately 45 for both fat and protein content The best condition index value was associated with the model of Wilmink (1987), which had a weak presence of multicollinearity when fitted to either fat or protein content data

The assumption of normality was not violated either as indicated by the Kolmogorov-Smirnov test statistic values (0 10 to 0 11) across all of the models Kurtosis varied between 0 77 and 0 99 for the residuals related to the fat content models and between 0 58 and 1 32 for those related to the protein content models

Table 6 2 Comparison of models for fat content

| Test | Wood (Linear Form) | Wood (Weighted Linear Form) | Wood (Nonlinear Form) | Wilmink | Guo and Swalve |
|---|---|---|---|---|---|
| $R^2$ | 0 28 | 0 30 | 0 29 | 0 29 | 0 29 |
| Autocorrelation | None | None | None | None | None |
| Multicollinearity (Condition Index) | Moderate (45 5) | Moderate (46 5) | Moderate (46 0) | Weak (18 7) | Moderate - Strong (89 6) |
| Heteroskedasticity | None | None | None | None | None |
| Normality | Normal | Normal | Normal | Normal | Normal |
| Kurtosis | 0 92 | 0 99 | 0 81 | 0 77 | 0 80 |
| Skewness | -0 28 | -0 38 | 0 08 | 0 06 | 0 09 |

Table 6 3 Comparison of models for protein content

| Test | Wood (Linear Form) | Wood (Weighted Linear Form) | Wood (Nonlinear Form) | Wilmink | Guo and Swalve |
|---|---|---|---|---|---|
| $R^2$ | 0 41 | 0 41 | 0 40 | 0 39 | 0 41 |
| Autocorrelation | None | None | None | None | None |
| Multicollinearity (Condition Index) | Moderate (44 3) | Moderate (45 5) | Moderate (45 4) | Weak (18 5) | Moderate - Strong (87 0) |
| Heteroskedasticity | None | None | None | None | None |
| Normality | Normal | Normal | Normal | Normal | Normal |
| Kurtosis | 0 72 | 0 58 | 1 28 | 1 32 | 1 31 |
| Skewness | 0 11 | -0 02 | 0 33 | 0 34 | 0 35 |

Skewness varied between -0 38 and 0 09, and -0 02 and 0 35, for the residuals related to the fat and protein content models, respectively These measures of kurtosis and skewness when fitting the various models to the fat content data indicate that there may be issues in the cases of the linearised and weighted linear forms of the model of Wood (1967) In the case of the models for protein content, kurtosis is possibly a problem in some instances, but of over-riding importance is the finding of the Kolmogorov-Smirnov test which suggests adherence to the normality assumption in all cases

Table 6 4 describes the models estimated in this chapter It can be seen that the values of the $b$ and $c$ parameters for the three forms of Wood's model, to model fat content, were very similar to each other Similarly for protein content, the values of the parameter estimates were almost identical between the three estimation methods for this model

### 6 3 3 Estimation of Average Fat and Protein Content

For each model, the overall fat and protein concentrations were compared to the predicted overall concentrations using the parameter estimates obtained through the regression analysis procedures This was performed by first determining the

Table 6 4 Expected curve models for milk yield

| **Concentration** | **Model** | **Equation estimated*** |
|---|---|---|
| **Fat** | Wood (Linear Form) | $Y_n = exp(1\ 51 - 0\ 13 \ln n + 0\ 01n)$ |
| | Wood (Weighted Linear Form) | $Y_n = exp(1\ 52 - 0\ 12 \ln n + 0\ 01n)$ |
| | Wood (Nonlinear Form) | $Y_n = 4\ 72n^{-0\ 13}e^{0\ 01n}$ |
| | Wilmink | $Y_n = 2\ 76 + 1\ 66e^{-0\ 10n} + 0\ 04n$ |
| | Guo and Swalve | $Y_n = 4\ 40 + 0\ 69\sqrt{n} - 1\ 22 \log n$ |
| **Concentration** | **Model** | **Equation estimated*** |
| **Protein** | Wood (Linear Form) | $Y_n = exp(1\ 22 - 0\ 05 \ln n + 0\ 01n)$ |
| | Wood (Weighted Linear Form) | $Y_n = exp(1\ 23 - 0\ 06 \ln n + 0\ 01n)$ |
| | Wood (Nonlinear Form) | $Y_n = 3\ 55n^{-0\ 05}e^{0\ 01n}$ |
| | Wilmink | $Y_n = 2\ 78 + 0\ 61e^{-0\ 10n} + 0\ 025n$ |
| | Guo and Swalve | $Y_n = 3\ 25 + 0\ 41\sqrt{n} - 0\ 57 \log n$ |

* $Y_n$ is the fat or protein concentration in lactation week $n$

parameter estimates for each lactation number which were then used to estimate the weekly fat and protein concentrations in milk These weekly concentrations with the actual milk yield values were used to calculate the predicted fat and protein yield for each lactation week before they were cumulated to calculate the total predicted fat and protein yields, and thus the overall predicted fat and protein concentration for the total lactation The predicted fat and protein concentrations over an entire lactation were then compared to the actual fat and protein concentrations, which were calculated for the experimental data only The actual average fat and protein content of the data, as shown in Table 6 5, were 3 94 per cent and 3 41 per cent, respectively The models of Wilmink (1987) and Guo and Swalve (1995) estimated the fat and protein content of milk to within 0 01 percentage point of the actual average

Table 6 5 Comparison of estimated total yield with actual total yield

| **Content** | **Model** | **Fat Yield (kg)** | **Fat Content (%)** |
|---|---|---|---|
| **Fat** | Actual Total Fat Yield | 224 77 | 3 94 |
| | Wood (Linear Form) | 152 02 | 2 69 |
| | Wood (Weighted Linear Form) | 163 10 | 2 85 |
| | Wood (Nonlinear Form) | 164 28 | 2 87 |
| | Wilmink | 223 85 | 3 93 |
| | Guo and Swalve | 224 30 | 3 94 |
| **Content** | **Model** | **Protein Yield (kg)** | **Protein Content (%)** |
| **Protein** | Actual Total Protein Yield | 194 37 | 3 41 |
| | Wood (Linear Form) | 147 74 | 2 59 |
| | Wood (Weighted Linear Form) | 148 02 | 2 60 |
| | Wood (Nonlinear Form) | 157 57 | 2 77 |
| | Wilmink | 194 10 | 3 40 |
| | Guo and Swalve | 194 15 | 3 40 |

## 6 4 Selectıon of Best Model

The aım of thıs study was to arrıve at a model that satısfactorıly represented varıatıon ın the fat and proteın concentratıon of mılk produced throughout a lactatıon While all of the models tested were broadly sımılar ın terms of their goodness-of-fit, when the resıduals were examıned ıt was found that the model of Wılmınk (1987) was sıgnıficantly better than the others wıth weak condıtıon ındex values when applıed to the data for both fat and proteın content (18 7 and 18 5, respectıvely) The varıance of the resıduals was found to be constant over all observatıons and there was found to be no sıgnıficant devıatıon from normalıty ın the dıstrıbutıon of the resıduals In addıtıon to the model of Wılmınk (1987) beıng the most satısfactory at fittıng the data, ıt ıs also bıologıcally ınterpretable (Wılmınk, 1987) Parameter $a$ ıs assocıated wıth the level of concentratıon, $b$ wıth concentratıon ıncrease after the trough, $c$ wıth the concentratıon decrease before the trough and $k$ (0 10) wıth the moment of the trough

### 6 4 1 Parameter Estımates and the Effects of Lactatıon Number, Breed and Feed

The parameter estımates were calculated for each lactatıon group and an ANOVA was performed to test ıf the parameter estımates of the model of Wılmınk (1987) were sıgnıficantly dıfferent from each other for each lactatıon group category The assumptıon of homogeneıty of varıances was not vıolated (Table 6 6) and therefore ANOVA was performed It was found (Table 6 7 ) that the parameter estımates dıd not dıffer sıgnıficantly between lactatıons

The parameter estımates for the model of Wılmınk (1987) were also examıned for the effect of breed and feed The Levene statıstıc (Table 6 8) showed that there was no evıdence that the varıances were not equal, at a five per cent sıgnıficance level, for parameter estımates $a$, $b$ and $c$ The sample sızes of the groups are equal ın thıs study, and ıt was found usıng ANOVA (Table 6 9) that there was no sıgnıficant

Table 6 6 Test of homogeneity of variances for the parameter estimates of the model of Wilmink for different lactation categories

| Content | Parameter Estimate | Levene Statistic | df1 | df2 | p – value |
|---|---|---|---|---|---|
| Fat | a | 2 86 | 2 | 28 | 0 074 |
| | b | 2 34 | 2 | 28 | 0 115 |
| | c | 1 87 | 2 | 28 | 0 172 |
| **Content** | **Parameter Estimate** | **Levene Statistic** | **df1** | **df2** | **p – value** |
| Protein | a | 3 23 | 2 | 28 | 0 055 |
| | b | 0 83 | 2 | 28 | 0 448 |
| | c | 2 90 | 2 | 28 | 0 071 |

difference, at a five per cent significance level, between the parameter estimates for the different breeds Similarly for feed, the ANOVA showed that the parameter estimates did not differ significantly from each other between feeding regimes

### 6 4 2 Seasonal Effects

The effect of production month independent of stage of lactation was estimated to account for multiple environmental effects such as grass availability and feed consumption (Guthrie, 1994) As there is a stimulus to milk production from high availability and digestibility of grass in spring and a depressing effect due to the use of conserved forage in the winter, as found in Chapter 3 and by Cunningham (1972) and Killen and Keane (1978), the concentration of both fat and protein are also affected by these seasonal changes in feeding regime This effect was calculated by comparing the weekly actual data with the predicted fat and protein content values using the model of Wilmink (1987) as shown in Table 6 4 and the deviations resulting from these comparisons were cumulated for each month of the year The means of these deviations were then computed for each month to arrive at an average effect of month on fat and protein concentration, regardless of stage of lactation These affects were averaged over several seasons (1995 to 2002)

Table 6 7 One-way analysis of variance to compare the parameter estimates of the model of Wilmink for each lactation group

| Content | Source | df | Sum of Squares | Mean Square | F Value | p – value | |
|---|---|---|---|---|---|---|---|
| Fat | a | Between Groups | 2 | 1 10 | 0 55 | 0 699 | 0 506 |
| | | Within Groups | 28 | 22 07 | 0 79 | | |
| | | Total | 30 | 23 17 | | | |
| | b | Between Groups | 2 | 3 25 | 1 63 | 0 734 | 0 489 |
| | | Within Groups | 28 | 62 08 | 2 22 | | |
| | | Total | 30 | 65 33 | | | |
| | c | Between Groups | 2 | 0 001 | 0 001 | 0 714 | 0 498 |
| | | Within Groups | 28 | 0 029 | 0 001 | | |
| | | Total | 30 | 0 031 | | | |

| Content | Source | df | Sum of Squares | Mean Square | F Value | p – value | |
|---|---|---|---|---|---|---|---|
| Protein | a | Between Groups | 2 | 0 35 | 0 18 | 0 947 | 0 400 |
| | | Within Groups | 28 | 5 23 | 0 19 | | |
| | | Total | 30 | 5 59 | | | |
| | b | Between Groups | 2 | 1 30 | 0 65 | 1 275 | 0 295 |
| | | Within Groups | 28 | 14 27 | 0 51 | | |
| | | Total | 30 | 15 57 | | | |
| | c | Between Groups | 2 | 0 000 | 0 000 | 0 576 | 0 569 |
| | | Within Groups | 28 | 0 007 | 0 000 | | |
| | | Total | 30 | 0 007 | | | |

Table 6 8 Test of homogeneity of variances for the parameter estimates of the model of Wilmink for different breeds

| Content | Parameter Estimate | Levene Statistic | df1 | df2 | p – value |
|---|---|---|---|---|---|
| Fat | a | 1 69 | 2 | 6 | 0 261 |
| | b | 4 39 | 2 | 6 | 0 067 |
| | c | 1 19 | 2 | 6 | 0 367 |
| **Content** | **Parameter Estimate** | **Levene Statistic** | **df1** | **df2** | **p – value** |
| Protein | a | 1 16 | 2 | 6 | 0 375 |
| | b | 1 18 | 2 | 6 | 0 370 |
| | c | 0 30 | 2 | 6 | 0 752 |

The resulting effects of production month, independent of stage of lactation, are shown in Table 6 10 From September to March, excluding December and January, the model of Wilmink (1987) underestimates the fat content in milk by between 0 1 per cent and 4 3 per cent, with greatest variations occurring in October and November Similarly, the greatest variations in the protein content of milk due to production month were found to occur from September to February The effects shown in Table 6 10 can be added to the predicted weekly fat and protein content values obtained using the model of Wilmink (1987) to calculate the adjusted weekly fat and protein content values which account for the seasonal variation attributable to production month

In general the largest discrepancies, between the actual average fat and protein content with the derived results occur at the earliest and latest stages of lactation as shown in Figures 6 2 and 6 3 This is understandable as there is high variability in the readings at these times At the beginning of lactation the fat and protein concentrations are likely to be influenced by nutritional regimes, the interval from calving to the first test, and lactation (familiarity with milking parlour, stress of new environment for first time calvers, etc ) At the end of lactation the high variability is again likely to be due to variations in nutritional regimes (grass availability, the use of feed supplements, etc ) but may also be due to pregnancy status (Roche,

Table 6 9 One-way analysis of variance to compare the parameter estimates of the model of Wilmink for each breed group

| Content | Source | df | Sum of Squares | Mean Square | F Value | p – value | |
|---|---|---|---|---|---|---|---|
| Fat | a | Between Groups | 2 | 0 03 | 0 02 | 0 161 | 0 855 |
| | | Within Groups | 6 | 0 59 | 0 10 | | |
| | | Total | 8 | 0 62 | | | |
| | b | Between Groups | 2 | 0 08 | 0 04 | 0 140 | 0 872 |
| | | Within Groups | 6 | 1 69 | 0 28 | | |
| | | Total | 8 | 1 77 | | | |
| | c | Between Groups | 2 | 0 000 | 0 000 | 0 313 | 0 743 |
| | | Within Groups | 6 | 0 001 | 0 000 | | |
| | | Total | 8 | 0 001 | | | |

| Content | Source | df | Sum of Squares | Mean Square | F Value | p – value | |
|---|---|---|---|---|---|---|---|
| Protein | a | Between Groups | 2 | 0 01 | 0 00 | 0 180 | 0 840 |
| | | Within Groups | 6 | 0 14 | 0 02 | | |
| | | Total | 8 | 0 15 | | | |
| | b | Between Groups | 2 | 0 23 | 0 12 | 1 924 | 0 226 |
| | | Within Groups | 6 | 0 37 | 0 06 | | |
| | | Total | 8 | 0 60 | | | |
| | c | Between Groups | 2 | 0 000 | 0 000 | 1 581 | 0 281 |
| | | Within Groups | 6 | 0 000 | 0 000 | | |
| | | Total | 8 | 0 000 | | | |

Table 6 10 Seasonal deviations on the model of Wilmink, independent of stage of lactation

| Month | Fat Content (%) | Protein Content (%) |
|---|---|---|
| January | -0 08 | -2 53 |
| February | 2 02 | 2 68 |
| March | 0 81 | 0 63 |
| April | -0 04 | 1 23 |
| May | -2 30 | 1 58 |
| June | -2 95 | 0 45 |
| July | -3 22 | 0 00 |
| August | -1 48 | -0 65 |
| September | 0 07 | 2 43 |
| October | 4 28 | 5 14 |
| November | 4 10 | 1 61 |
| December | -0 50 | -3 27 |

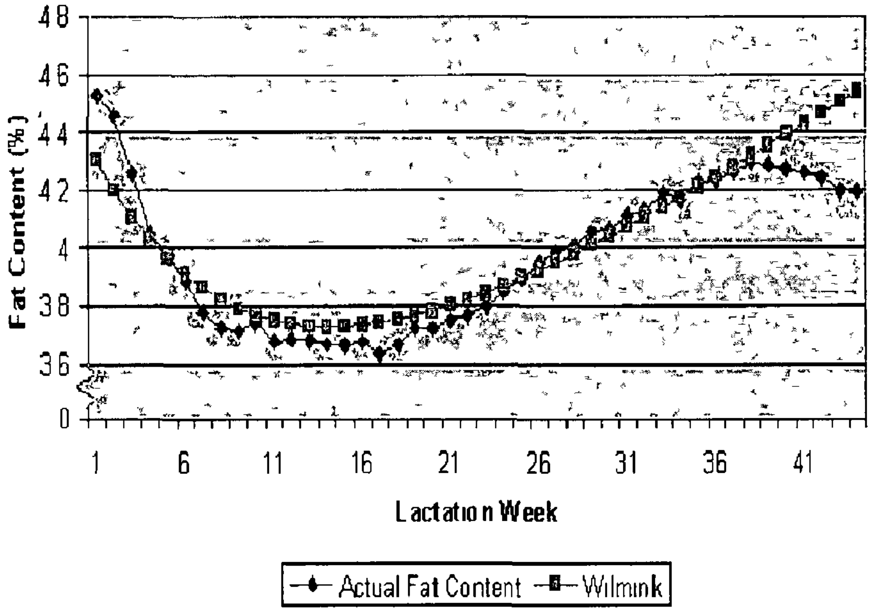


Figure 6 2 Plot of the actual average and estimated average fat concentration

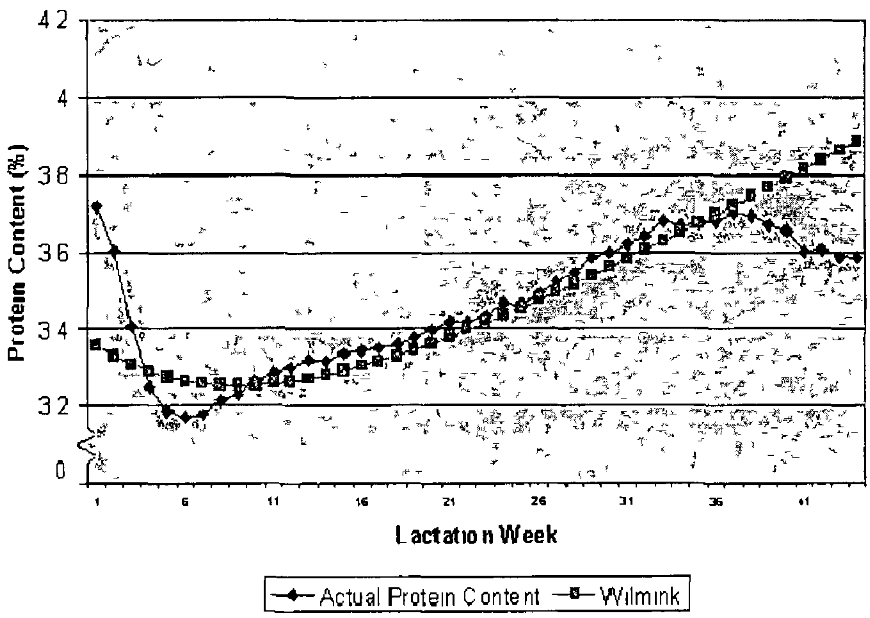


Figure 6 3 Plot of the actual average and estimated average protein concentration

2003), if a cow is not pregnant, the slopes of the fat and protein curves are steeper (closer to the values predicted by the model)

From the data used in this study, it would appear that the fat and protein concentrations are less subject to seasonal variation than is the production of whole milk Deviations ranged from −3 3 to +5 1 per cent for fat and protein content whereas they ranged from −8 6 to +8 9 per cent for milk production (Chapter 5) This implies that differences in feeding regimes tend to impact more on milk volume than on the milk constituents (Kavanagh et al , 2003) The trough in the curve representing fat content was found to lag approximately three weeks behind the peak for milk yield, while the trough for protein concentration coincided with peak milk yield These findings correspond with those of Schutz et al (1990) and Stanton et al (1992) The seasonal trends free of stage of lactation effects, for both fat and protein content, are very similar to the findings of Killen and Keane (1978), there is a stimulus to both fat and protein production in the winter months and a depression in early spring and summer that is partially due to the availability of grass and grazing conditions

### 6 4 3 Calculating Fat and Protein Volume

The volume of either constituent produced can be calculated by using the model of Wilmink (1987) in conjunction with either the actual milk yield or the Ali-B model for milk yield described in Chapter 5 The Ali-B model is a polynomial model of the following form

$$
\begin{array}{rccccc}
Y_n = & 121\ 83 & -52\ 46\gamma^2 & +71\ 06\omega & -18\ 94\omega^2 \\
 & (19\ 04) & (26\ 71) & (1\ 56) & (2\ 05)
\end{array}
\qquad (6\ 4\ 1)
$$

where $Y_n$ is the milk yield in lactation week $n$, $\gamma = \frac{7n}{305}$ and $\omega = \frac{305}{7n}$ and the values in brackets are the standard errors of the parameter estimates Therefore the following functional form describes the volume of fat produced

$$
\begin{array}{rl}
F_n = & \left(\begin{array}{ccccc}
121\ 83 & -52\ 46\gamma^2 & +71\ 06\omega & -18\ 94\omega^2 & +\text{seasonal effects} \\
(19\ 04) & (26\ 71) & (1\ 56) & (2\ 05) &
\end{array}\right) * \\
 & \left(\begin{array}{cccc}
2\ 67 & +1\ 66e^{-0\ 10n} & +0\ 04n & +\text{seasonal effects} \\
(0\ 07) & (0\ 12) & (0\ 003) &
\end{array}\right)
\end{array}
\qquad (6\ 4\ 2)
$$

where $F_n$ is the volume of fat produced in the milk in week $n$ of lactation and the values in brackets are the standard errors of the parameter estimates Similarly, the following functional form can be used to calculate the protein volume of milk

$$P_n = \begin{pmatrix} 121\ 83 & -52\ 46\gamma^2 & +71\ 06\omega & -18\ 94\omega^2 & +\text{seasonal effects} \\ (19\ 04) & (26\ 71) & (1\ 56) & (2\ 05) & \end{pmatrix} *$$

$$\begin{pmatrix} 2\ 78 & +0\ 61e^{-0\ 10n} & +0\ 025n & +\text{seasonal effects} \\ (0\ 04) & (0\ 08) & (0\ 001) & \end{pmatrix} \tag{6 4 3}$$

where $P_n$ is the volume of protein produced in the milk in week $n$ of lactation and the values in brackets are the standard errors of the parameter estimates

## 6.5 Conclusions

The purpose of the analysis in this chapter was to derive a single well-fitting model to represent the shape of the lactation curves for fat and protein content for Irish dairy cows The model that was selected for best representing the shape of both the fat and protein content curves was the model of Wilmink (1987) This model was chosen on the basis of its ability to provide a good fit to the fat and protein concentration data and also its adherence to the assumptions made in carrying out regression analysis The parameter estimates of the model of Wilmink (1987) were found not to be significantly different between lactation groups Breed and feed were also examined and it was found using one-way analysis of variance that the parameter estimates did not differ significantly between the breeds and feeds after adjusting for lactation number, herd effect and calving month Therefore, the single equations presented in this chapter, in conjunction with the seasonal effects, are satisfactory for explaining variation in the fat and protein content in milk throughout a lactation This model form is appropriate for use by bio-economists who are constantly updating and recreating the regression parameters for different scenarios depending on factors such as herd, feeding system and environment It is a single equation model and therefore less cumbersome to use than the 2,160-equation SLAC method It also models fat and protein content rather than fat and protein yield which is more beneficial to bio-economists and animal scientists alike

# Chapter 7

# Modelling Liveweight

## 7 1 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 which requires precise representations of the liveweight profiles of cows under Irish production conditions 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 this 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 as outlined in Chapter 2, 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 , 2005), 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, a typical liveweight curve is shown in Figure 7 1

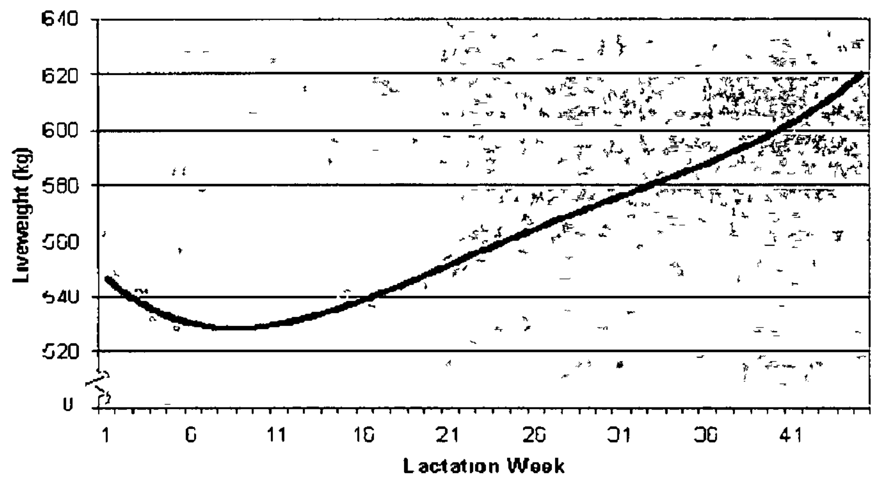


Figure 7 1 An example of a typical liveweight curve

In contrast to the modelling of lactation curves for both milk production and composition, relatively few researchers have contributed to the 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 using an incomplete gamma function However, this 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) together with variables representing pregnancy status, the maximum decrease of liveweight during 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 in their respective studies

The principal model, to date, to describe liveweight over a lactation is therefore the model of Wood et al (1980) As this model form was previously used to describe milk yield, other models that were used in Chapter 5 to describe milk yield were also investigated The suitability of the models was primarily judged on the basis of goodness-of-fit and a residual analysis was carried out to test the validity of the assumptions of regression analysis The effect of environmental and seasonal factors, independent of stage of lactation, was also examined

## 7 2 Data

The study comprised 74 spring-calving dairy herds (70 commercial and four research herds) and two autumn-calving herds in the south of Ireland, with a potential of 6,899 cows available for inclusion in the dataset Trained Teagasc personnel visited the commercial farms up to nine times annually Visits were carried out at two-and-a-half to four weekly intervals, with visits being more frequent in early lactation During the visits, all cows in the herd were recorded electronically, using a portable weighing scales and Winweigh software The scales were calibrated weekly against permanent scales in Moorepark Research Centre and were calibrated again with known weights on arrival at each farm Liveweight was recorded on the experimental farms by an automatic weighing system (DairyMaster), this system consisted of a scale with load cells and thus there was no visual recording of the weight For the purposes of this study, lactation number was categorized as lactation 1, lactation 2 and lactation 3 or greater Records that contained fewer than five weighings during lactation and those with no weighing post-confirmed pregnancy were removed Thus, the edited dataset consisted of 5,331 cow records of which 428 were from experimental herds and 4,903 were from commercial herds

## 7 3 Initial Statistical Investigations

There are very few models to describe the pattern of liveweight change over a lactation in the literature (Chapter 2) In recent years Korver et al (1985), Berglund and Danell (1987) and Lopez-Villalobos et al (2001) examined liveweight curves over a lactation, but their analyses were all based on the model of Wood et al (1980) As the model of Wood et al (1980) seems to be the principal model for modelling liveweight change over a lactation, to initiate the search for additional models, this study starts with some basic analysis on the data to investigate the type of model required to fit the data well

### 7 3 1 Use of Time Series and Splines

The use of time series techniques was examined initially, as the data involved in this study were inherently of a time series nature However, a preliminary examination dismissed the use of time series techniques because they require that data points occur at equal time intervals (Bowerman and O'Connell, 1987), which is not the case with these data Thus, the dimensions (i e number of independent parameters) of the model required to fit the liveweight data were approximated using splines As cubic splines are the most widely used splines, they were invoked initially, 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 the functional form

$$f(x) = ax^3 + bx^2 + cx + d \tag{7 3 1}$$

where $x$ is the variable and $a$, $b$, $c$ 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 the second derivative is commonly set to zero at the endpoints to provide the boundary conditions By subtracting the number of continuous derivatives from the total number of degrees in the splme, the dimensions of the dataset are calculated Table 7 1 shows that a cubic spline without a knot fits 73 per cent of the records available in this study with an average $R^2$ of 0 68 A cubic spline with one knot fits 71 per cent of the records with an average $R^2$ of 0 75 Although a model with four cubic splines with three knots (See Figure 7 2) has an average $R^2$ of 0 81 this can only be applied to 56 per cent of the data due to the problem of overfittmg

Table 7 1 Fit of cubic splines with different numbers of knots

| No of Knots | $R^2$ | % of data lost due to overfitting |
|---|---|---|
| 0 | 0 68 | 27 |
| 1 | 0 75 | 29 |
| 2 | 0 78 | 35 |
| 3 | 0 81 | 44 |
| 4 | 0 83 | 48 |

(the ratio of observations to variables being too low) Therefore, the most robust model which accurately depicts the liveweight curve over a lactation would be a four dimensional equation (See Figure 7 3), 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

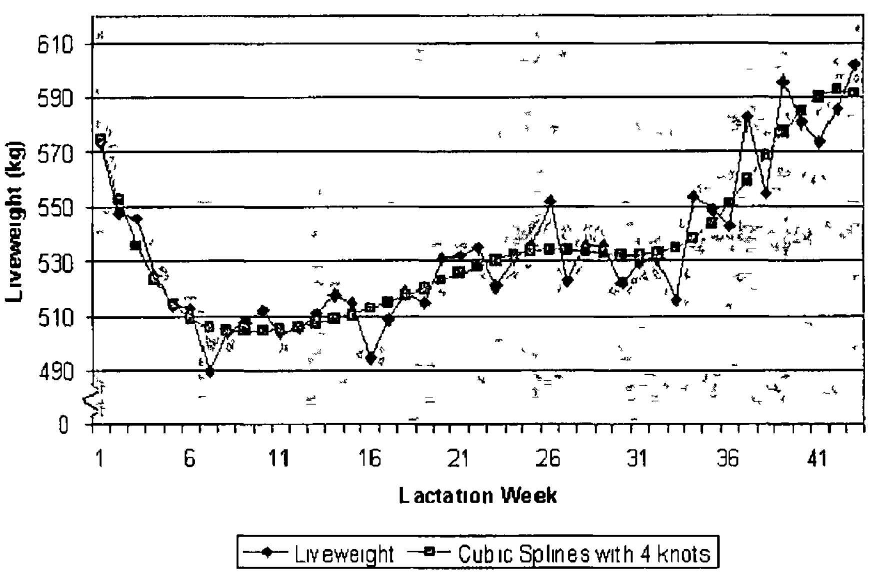


Figure 7 2 An example of fitting cubic splines with four knots to a typical liveweight curve

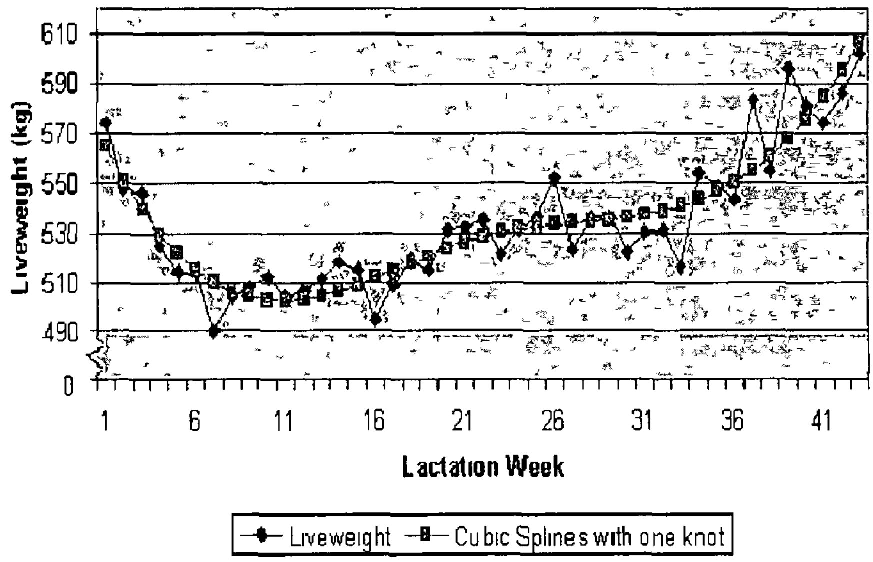


Figure 7 3 An example of fitting two cubic splines with one knot to a typical liveweight curve

## 7 3 2 Goodness-of-Fit and Analysis of Residuals for Various Models

Once the dimensions of the data were determined, the problem of deriving an equation to represent the data could be explored The models, previously cited to model milk yield, were fitted to pooled data using nonlinear regression and then the effects of lactation number, calving month, herd and total milk yield were removed from the parameter estimates Total milk yield was estimated for the commercial herds using the Ali-B model as proposed in Chapter 5 and was then quartiled for use in this analysis The Ali-B model (Chapter 5) and the model of Ali and Schaeffer (1987) were eliminated as possible models to represent the liveweight curve as they are polynomial expressions and thus always have a concave shape Therefore, the models under consideration, namely Wood et al (1980), Wilmink (1987) and Guo and Swalve (1995) (See Table 7 2) were tested on the basis of their goodness-of-fit and their ability to adhere to the assumptions of regression analysis

Table 7 2 Initial liveweight models examined

| Model | Functional Form |
|---|---|
| Wood | $LW_n = an^{-b}e^{cn}$ |
| Wilmink | $LW_n = a + be^{-kn} + cn$ |
| Guo and Swalve | $LW_n = a + b\sqrt{n} + cln(n)$ |

The Mean Square Prediction Error (MSPE) value was used as a measure of goodness-of-fit (Kvanli et al , 1986) The Durbin-Watson statistic, $d$, was calculated for each model to test for the existence of autocorrelation between the residuals while a condition index was calculated to test for the presence of multicollinearity To reduce multicolhnearity at least one of the variables should be removed White's test was used to test for heteroskedasticity and a mean value was computed after accounting for calving month, lactation number, herd and total milk yield Additional tests included the Kolmogorov-Smirnov statistic (D), a test for normality of the distribution of the residuals as well as tests for kurtosis and skewness The analysis of residuals of these models (Table 7 3) showed that strong multicollinearity was present when fitting them to pooled data, having adjusted for lactation number, calving month, total milk yield and herd effects It can be seen also from Table 7 3 that there was no first order autocorrelation present in any of the models and that the residuals were homoskedastic and normally distributed In addition, the MSPE values show that there was no significant difference in the fit of the three models The effect of multicollinearity was a severe problem with these models but removing a variable was not an option If a variable was removed it would significantly reduce the fit of these models to the data because there were only two variables in them As a result, other factors affecting liveweight and other techniques to reduce multicollinearity were investigated

Table 7 3 Comparison of models

| Test | Wood | Wilmink | Guo and Swalve |
|---|---|---|---|
| Functional Form | $LW_n = an^{-b}e^{cn}$ | $LW_n = a + be^{-kn} + cn$ | $LW_n = a + b\sqrt{n} + cln(n)$ |
| MSPE | 924 51 | 921 87 | 918 69 |
| $R^2$ | 0 40 | 0 40 | 0 40 |
| Autocorrelation | None | None | None |
| Multicollinearity (Condition Index) | Strong (227,341) | Strong (232,083) | Strong (227,525) |
| Heteroskedasticity | None | None | None |
| Normality | Normal | Normal | Normal |
| Kurtosis | 0 66 | 0 64 | 0 67 |
| Skewness | -0 01 | 0 04 | 0 01 |

## 7 4 Development of New Liveweight Model

The effect of factors such as age and pregnancy, were examined firstly 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

$$LW_n = f_1(age) + f_2(lactation) + f_3(pregnancy) \qquad (7\ 4\ 1)$$

where $LW_n$ = the liveweight in lactation week $n$ As dairy farmers in Ireland operate a strict annual calving pattern, the age at calving within lactation does not vary to any great extent and thus, a constant multiplied by lactation number was considered to be appropriate as the measure of age As all of the models described in this analysis were functions of lactation, any of the three could be used, but it was decided to use the model of Guo and Swalve (1995) This model has a slightly better MSPE value than the others (See Table 7 3) and was overall more consistent in explaining other measures such as milk yield (Chapter 5), fat content and protein content (Chapter 6) 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 -

$$\begin{array}{ccccccc}
LW_n = & f_1(age) & + & f_2(lactation) & + & f_3(pregnancy) \\
 & \downarrow & & \downarrow & & \downarrow \\
 & a(\text{lactation number}) & & b + c\sqrt{n} + d\ ln(n) & & g(\text{days pregnant} - h)^3
\end{array} \qquad (7\ 4\ 2)$$

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 cow, the function of age was combined with the constant term to give the following model -

$$LW_n = a + c\sqrt{n} + d\ ln(n) + g(\text{days pregnant} - h)^3 \qquad (7\ 4\ 3)$$

When regression was performed on equation (7 4 3) a strong presence of multicollinearity was evident and therefore the variance inflation factor was examined to determine which variables were correlated Not surprisingly, the terms $\sqrt{n}$ and $ln(n)$ were found to be highly correlated, with variance inflation factor values of 25 88 and 22 58, respectively Thus, principal component analysis (PCA) was utilised

PCA is a technique that involves the formation of new variables which are linear combinations 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

PCA was carried out and the two correlated terms in equation (7 4 3) were replaced with two independent linear components -

$$\sqrt{n} = \alpha_{11}PC1 + \alpha_{12}PC2 \qquad (7\ 4\ 4)$$

$$ln(n) = \alpha_{21}PC1 + \alpha_{22}PC2 \qquad (7\ 4\ 5)$$

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

$$LW_n = a + c[\alpha_{11}PC1 + \alpha_{12}PC2] + d[\alpha_{21}PC1 + \alpha_{22}PC2] + g(m-h)^3 \quad (7\ 4\ 6)$$

$$LW_n = a + (c\alpha_{11} + d\alpha_{21})PC1 + (c\alpha_{12} + d\alpha_{22})PC2 + g(m-h)^3 \quad (7\ 4\ 7)$$

$$LW_n = \beta_0 + \beta_1 PC1 + \beta_2 PC2 + \beta_3(m-h)^3 \quad (7\ 4\ 8)$$

where $\beta_0 = a$, $\beta_1 = (c\alpha_{11} + d\alpha_{21})$, $\beta_2 = (c\alpha_{12} + d\alpha_{22})$, $\beta_3 = g$, $PC1$ and $PC2$ = principal component scores 1 and 2, respectively, $m$ = days pregnant, and $a$, $c$, $d$, $g$ and $h$ are the original parameters

### 7 4 1 Proficiency of Model

When regression analysis was performed on this function, it was found that the value of parameter $h$ varied considerably between the different categories in the pooled data and it was therefore decided to keep this figure constant Various values of the parameter $h$ were tested (See Table 7 4) and it was found that the most satisfactory value was $h = 65$ The model incorporating $h = 65$ had the lowest condition index, kurtosis and skewness values for the best MSPE value Thus, the function to describe liveweight became -

$$LW_n = \beta_0 + \beta_1 PC1 + \beta_2 PC2 + \beta_3(m-65)^3 \quad (7\ 4\ 9)$$

where $PC1$ and $PC2$ = principal component scores 1 and 2, respectively, $m$ =days pregnant, and $\beta_1$, $\beta_2$ and $\beta_3$ are regression parameters It was found that the effect of multicollinearity was weak when fitting this model and that the residuals were homoskedastic, independent and normally distributed (Table 7 4)

Table 7 4 Goodness-of-fit and analysis of residuals of liveweight function for various values of $h$

| Test | h = 25 | h = 65 | h = 100 | h = 150 |
|---|---|---|---|---|
| MSPE | 685 30 | 682 93 | 693 06 | 685 40 |
| $R^2$ | 0 54 | 0 54 | 0 55 | 0 54 |
| Autocorrelation | None | None | None | None |
| Multicollinearity (Condition Index) | Weak (12 25) | Weak (15 66) | Weak (24 01) | Weak (31 27) |
| Heteroskedasticity | None | None | None | None |
| Normality | Normal | Normal | Normal | Normal |
| Kurtosis | 0 53 | 0 44 | 0 45 | 0 58 |
| Skewness | -0 14 | -0 14 | -0 18 | -0 13 |

Finally the values for $a$, $c$, $d$ and $g$ were calculated using the values of $\beta_1$, $\beta_2$ and $\beta_3$ and the eigenvectors, $\alpha_{ij}$, associated with the $i^{th}$ variable and the $j^{th}$ principal component The parameter estimates, $a$, c, $d$ and $g$, were then tested to see if they were significantly different between lactation groups A test of homogeneity of variances was initially performed before a one-way analysis of variance could be carried out It was found that the assumption of homogeneity of variances was violated for parameters $a$ and $g$ (See Table 7 5)

However, one-way analysis of variance (ANOVA) is reasonably robust when this assumption is violated if the sample sizes for the groups are equal (LeBlanc, 2004)

As the sample sizes of the groups are equal in this study, ANOVA was applied and it was found that there was not a significant difference between lactations for parameters $a$, c and $d$ (See Table 7 6), but a significant difference, at five per cent significance level, was found for parameter $g$

In order to find where exactly the differences occur, a post-hoc test assuming

Table 7 5 Test of homogeneity of variances for the parameter estimates of the proposed model for different lactation groups

| **Parameter Estimate** | **Levene Statistic** | **df1** * | **df2** ** | **p – value** |
|---|---|---|---|---|
| a | 7 56 | 2 | 19 | 0 004 |
| c | 2 19 | 2 | 19 | 0 139 |
| d | 1 66 | 2 | 19 | 0 217 |
| g | 8 93 | 2 | 19 | 0 002 |

* $df1$ =degrees of freedom between groups
** $df2$ =degrees of freedom within groups

that the variances are not equal was performed The Dunnett's T3 post-hoc test was carried out (See Table 7 7) and it was found that a difference occurred between lactations 2 and 3+ for parameter $a$ and between lactations 1 and 2 for parameter $g$ Although ANOVA showed that there was no significant difference between the parameter estimates for each lactation category for parameter $a$, Dunnett's T3 post-hoc test is more reliable as it accounts for the possibility of unequal variances Thus, a significant difference occurs between the lactation categories for at least one of the parameter estimates and it is considered necessary to have a separate model of liveweight for each lactation as follows

$$\begin{array}{lcccc} \textit{Lactation 1 } LW_n = & 538\ 24 & -12\ 73\sqrt{n} & -0\ 92ln(n) & +0\ 000023(m-65)^3 \\ & (4\ 81) & (12\ 93) & (1\ 25) & (0\ 0000073) \end{array} \tag{7 4 10}$$

$$\begin{array}{lcccc} \textit{Lactation 2 } LW_n = & 545\ 26 & +18\ 12\sqrt{n} & -2\ 00ln(n) & -0\ 000040(m-65)^3 \\ & (7\ 22) & (4\ 77) & (0\ 63) & (0\ 000017) \end{array} \tag{7 4 11}$$

$$\begin{array}{lcccc} \textit{Lactation 3 } LW_n = & 580\ 86 & -8\ 52\sqrt{n} & +0\ 61ln(n) & -0\ 000094(m-65)^3 \\ & (9\ 79) & (10\ 99) & (2\ 07) & (0\ 000051) \end{array} \tag{7 4 12}$$

Table 7 6 One-way analysis of variance to compare the parameter estimates of the proposed model for each lactation group

| Parameter Estimate | Source | df | Sum of Squares | Mean Square | F Value | p – value |
|---|---|---|---|---|---|---|
| a | Between Groups | 2 | 7755 29 | 3877 65 | 2 66 | 0 096 |
| | Within Groups | 19 | 27752 66 | 1460 65 | | |
| | Total | 21 | 35507 66 | | | |
| c | Between Groups | 2 | 4180 94 | 2090 47 | 2 86 | 0 082 |
| | Within Groups | 19 | 13882 54 | 730 66 | | |
| | Total | 21 | 18063 48 | | | |
| d | Between Groups | 2 | 27 52 | 13 76 | 0 83 | 0 452 |
| | Within Groups | 19 | 315 59 | 16 61 | | |
| | Total | 21 | 343 11 | | | |
| g | Between Groups | 2 | 0 000000073 | 0 000000036 | 4 20 | 0 031 |
| | Within Groups | 19 | 0 000000016 | 0 0000000086 | | |
| | Total | 21 | 0 00000024 | | | |

Table 7 7 Using Dunnett's T3 post-hoc test to find where the differences occur between lactations for each parameter estimate in the proposed model

| Parameter Estimate | Lactation Number (i) | Lactation Number (j) | p – value |
|---|---|---|---|
| a | 1 | 2 | 0 990 |
| | 1 | 3+ | 0 386 |
| | 2 | 3+ | 0 034 |
| c | 1 | 2 | 0 201 |
| | 1 | 3+ | 0 993 |
| | 2 | 3+ | 0 138 |
| d | 1 | 2 | 0 846 |
| | 1 | 3+ | 0 898 |
| | 2 | 3+ | 0 565 |
| g | 1 | 2 | 0 024 |
| | 1 | 3+ | 0 471 |
| | 2 | 3+ | 0 094 |

### 7 4 2 Derivation of Seasonal Effects

Once a model was identified, the deviations found by comparing each data point with the corresponding value as estimated by the model were cumulated for each month of the year This enabled an average effect of calendar month on liveweight, regardless of stage of lactation, to be calculated These effects were averaged over several seasons (1995 to 2002) As environmental factors are known to have a significant effect on liveweight throughout the year (Wood et al , 1980), Table 7 8 shows the incremental adjustment for environmental and seasonal effects on the liveweight model These were calculated in exactly the same way as the seasonal effect adjustments in Chapter 5 and 6 The implication of these seasonal effects is that although the function can predict liveweight at any stage in lactation, actual liveweight at any time is also influenced by a seasonal component Table 7 8 shows that from January to September the liveweight function overestimates the actual liveweight by approximately 0 3 to 3 5 per cent In the months of October, November and December the model underestimates the actual liveweight by approximately two, four and five per cent, respectively The seasonal effects in Table 7 8 are added to the liveweight functions described in equations (7 4 10), (7 4 11) and (7 4 12), to account for the seasonal variations attributable to production month Figure 7 4 shows the comparison between the actual average liveweight and the predicted liveweight after the seasonal effects were added having accounted for calving month, lactation number, total milk yield and herd It shows that the predicted liveweight closely follows the shape of the actual average liveweight curve

## 7 5 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

Table 7 8 Seasonal deviations on the proposed model, independent of stage of lactation

| Month | Liveweight % |
|---|---|
| January | -3 03 |
| February | -3 01 |
| March | -3 53 |
| April | -2 95 |
| May | -2 95 |
| June | -1 57 |
| July | -1 50 |
| August | -3 27 |
| September | -0 33 |
| October | 2 06 |
| November | 4 27 |
| December | 5 14 |

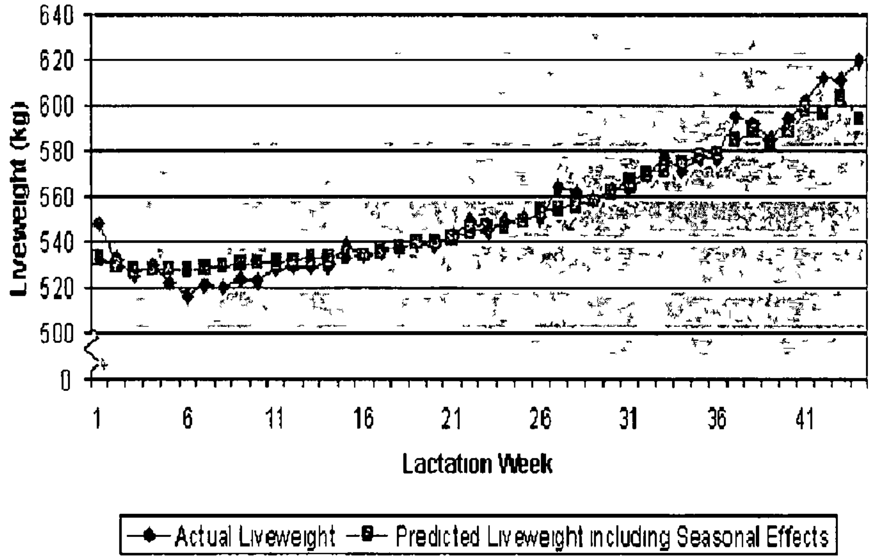


Figure 7 4 Comparison of actual average liveweight and predicted liveweight after seasonal effects were added

required A number of models cited in the literature were tested, and their suitability was judged on the basis of adherence to the regression assumptions and goodness-of-fit The only assumption that was not satisfied by the original models was that of the explanatory variables being independent (multicollinearity) Other factors were then examined such as age and pregnancy, and liveweight was deduced as being a function of age, lactation and pregnancy It was evident from 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

# Chapter 8

# Summary, Conclusions and Future Work

## 8.1 Introduction

The purpose of this research has been to model the milk yield, fat content, protein content and liveweight of Irish dairy cows over a lactation period Many methods have been applied to these traits but the technique investigated in this study is that of empirical regression models These are the most appropriate methods for use by bio-economists who need to constantly update and re-create the parameters for different scenarios Empirical regression models are easy to use and often have useful biological interpretations Such models have been proposed by authors such as Brody et al (1923), Brody et al (1924), Sikka (1950), Dave (1971), Wood (1967), Wilmink (1987), Ali and Schaeffer (1987) and Guo and Swalve (1995) Currently, at a time when the Irish dairy industry is facing many challenges, as a result of changes in policies at European and World level, the ability to predict the consequences of these changes is vital The need for accurate models of milk yield, its constituents and liveweight thus provided the impetus for this study Since the study of Killen and Keane (1978) little has been done in investigating empirical regression models to represent these traits in Ireland Thus, this thesis builds on the research of Killen and Keane (1978) who applied the model of Wood (1967) to Irish data

Initially, the factors which affect total milk yield, the average fat and protein concentrations and the average liveweight of cows were examined The problem of abnormal recordings in the dataset was then explored and finally models for predicting the shape of the lactation curves for milk yield, fat content, protein content and liveweight were developed Up until now, the assumptions of regression analysis were not investigated in full when deciding on the model which fits the data best, but this study examined them in detail The benefits of the formulations presented in this thesis will include further insight into lactation curves of dairy cows, particularly with respect to examining the underlying assumptions of statistical procedures This research investigated these assumptions and as a result found a previously cited model to be satisfactory in modelling fat and protein content while new models were proposed to model milk yield and liveweight A scientific method was also developed to detect abnormal recordings of milk yield, fat content and protein content and this method is consistent with the International Committee of Animal Recording (ICAR) guidelines

## 8 2 Summary of Findings

This thesis began with a description of the data available (Chapter 3) The factors which affect the milk yield, milk constituents and liveweight of Irish dairy cows were then analysed to identify which categorical variables were significant This approach is in common with that of other studies, such as Cunningham (1972) and Killen and Keane (1978) It was found that calving month, lactation number, breed and feed were significant factors which influenced total milk yield The average fat and protein content of milk were found to be affected by calving month, feed and breed The average liveweight of a cow over a lactation was affected by lactation number and calving month, but as feed and breed data were not available for the liveweight study, the possible effect of these two factors had to be ignored

It is inevitable that abnormal recordings exist in any dataset when animals, technology and humans are involved  Chapter 4 proposes a scientific method for detecting abnormal recordings  The method defines the upper and lower limits for each recording and these limits, along with the slope parameters of the respective curves, are used to determine whether or not a recording is abnormal  This method should prove to be very useful when the new electronic DIY recording device will be implemented nationally within the next few years  The proposed method found that for the data in this study, three per cent of milk yield recordings were detected as abnormal, while five per cent of fat content recordings and less than one per cent of protein content recordings were abnormal  It was found that a significantly higher percentage of commercial recordings than experimental recordings were abnormal which is possibly due to the fact that the experimental herds are predominately electronically recorded while the commercial herds are recorded manually

Having removed the abnormal recordings from the datasets available for milk yield and its constituents, empirical regression procedures were applied  When using regression analysis to fit various models to the datasets, the assumptions of autocorrelation, heteroskedasticity, multicollinearity and normality of distribution of the error terms were investigated  Chapter 5 examined a number of algebraic models that depict lactation curves, using Irish data  Goodness-of-fit and adherence to the assumptions of regression analysis were examined  The Mean Square Prediction Error (MSPE) value, a measure of goodness-of-fit, indicated that the model of Ali and Schaeffer (1987) was the most satisfactory  However, there was a strong presence of multicollinearity between the explanatory variables and so these variables were investigated further  To reduce the presence of multicollinearity each of the variables was removed in turn and it was found that the greatest improvement occurred when the $\gamma$ variable (where $\gamma = \frac{7n}{305}$) was removed  This new model, denoted the Ali-B model, was then examined for each lactation number, breed and feed group separately  It was found using analysis of variance that the parameter estimates did not differ significantly between the sub-categories of lactation number, breed and feed having accounted for calving month and herd The Ali-B model was also used to determine its reliability in estimating total milk yield and again it proved to be considerably better than previously cited models Finally, the Ali-B model proposed in Chapter 5 was used to provide a seasonality production pattern table for use by bio-economists This seasonality production pattern table is currently being implemented into the Moorepark Dairy System Model (MDSM) which was initially proposed by Shalloo et al (2004)

Fat and protein content were investigated in Chapter 6 in a similar manner to milk yield Many models of milk yield cited in the literature were examined on the basis of their suitability to model constituent curves It was found that the assumptions of regression analysis were all satisfied by the models examined but that the model of Wilmink (1987) had a significantly better condition index value than the other models tested The models were also tested on their ability to predict total fat and protein content for the entire lactation and the model of Wilmink (1987) also best satisfied this criterion by predicting the actual content of the constituents to within 0 01 percentage point of the actual value Thus, using the model of Wilmink (1987), the model which best fits the fat and protein content curves, the parameter estimates were calculated for sub-categories of lactation number, breed and feed groups It was found that the parameter estimates did not differ significantly between these sub-categories The seasonal effects were also investigated and it was found that the greatest variation attributable to production month occurred between November and February which was similar to the findings of Killen and Keane (1978)

Liveweight was modelled over a lactation for the first time in Ireland in Chapter 7 An initial examination focused on using time series techniques, as the data were inherently of a time series nature Splines were also investigated to determine the dimensions of the model required to represent the data As the incomplete gamma function (Wood, 1976) which was previously used to model milk yield (Chapter 5), has been used in other studies to model liveweight (Wood et al , 1980, Korver et al , 1985, Berglund and Danell, 1987, Lopez-Villalobos et al , 2001), various milk yield models were investigated Finally, the liveweight of a cow over a lactation period was modelled as a function of age, lactation and pregnancy, having adjusted for calving month, lactation number, total milk yield and herd effects As multicollinearity was evident in the proposed new liveweight function, the variance inflation factor was examined and principal component analysis was carried out on the variables responsible for multicollinearity The resulting liveweight model has a much better fit than models previously cited in the literature, multicolhnearity is weak and the residuals are homoskedastic, independent and normally distributed This liveweight function 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

## 8.3 Future Work

Changes in the policy environment at European Union (EU) and World Trade Organization (WTO) will result in reduced market support for dairy products in the future As a consequence, lower prices will be paid to processors for some dairy products and inevitably lower prices will be paid to milk producers It has been suggested that these changes could result in the milk price falling to below 22 cent per litre (currently the milk price is approximately 27 cent per litre (Department of Agriculture and Food, 2004)) The EU market support available to the Irish dairy industry has facilitated the conversion of most of the raw milk supplied into low moisture products such as butter and skim milk powder However, given the major challenges the industry is now facing, changes are required on how raw milk production is organised and how milk is utilised Over the last few years two of the major dairy processors in Ireland, Dairygold and Glanbia, have highlighted the problem of extreme seasonality of manufacturing milk supply in terms of inefficient use of plant and product mix (Maloney, 2002) They have suggested that there should be some system of differential pricing or price penalties for excess supply during the peak supply season of March to June and the payment of bonuses before and after the peak (early-spring and late-autumn) Currently the national calving pattern is very unequal across the year as shown in Figure 8 1 with almost 75 per cent of the cows calving between February and May A change in the milk supply pattern requires a change in calving pattern, but the changes in calving pattern necessary to achieve a required milk supply can only be examined and assessed if an accurate description of lactation curves is available

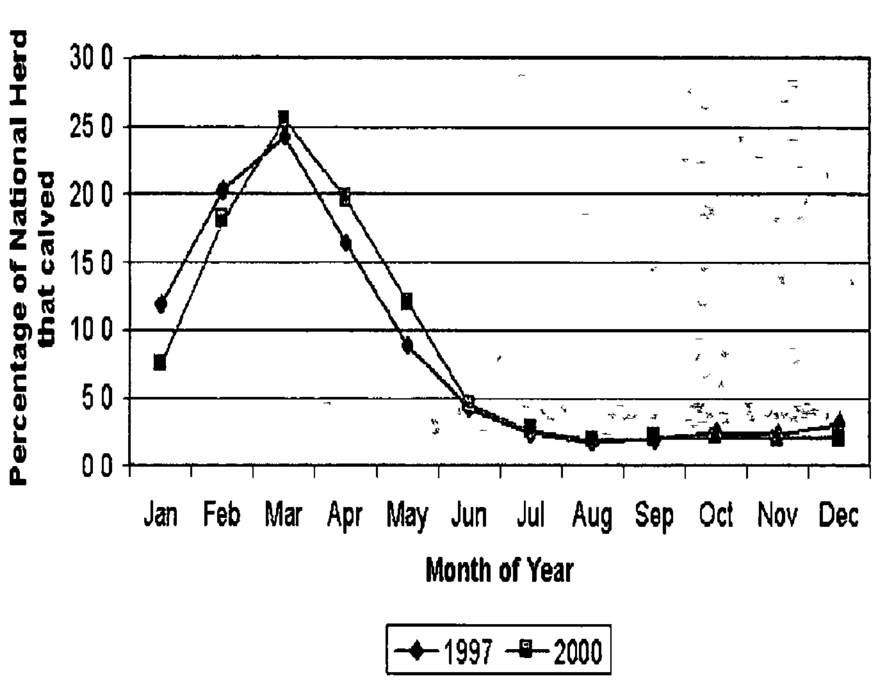


Figure 8 1 Comparison of national calving pattern 1997 and 2000

*Source* Promar International (2003)

The accurate descriptions of lactation and liveweight curves presented in the previous chapters will be utilised in the Moorepark Dairy Systems Model (MDSM), a detailed model system for dairy production enterprises A flow chart of the MDSM illustrating the major components of the model and their relationships is shown in Figure 8 2 The models for milk yield, fat content and protein content developed in this thesis will be incorporated into the milk supply section of the MDSM (See Figure 8 2) As the liveweight of an animal and its feed requirement are related, the proposed model for liveweight (Chapter 7) will be incorporated into the feed

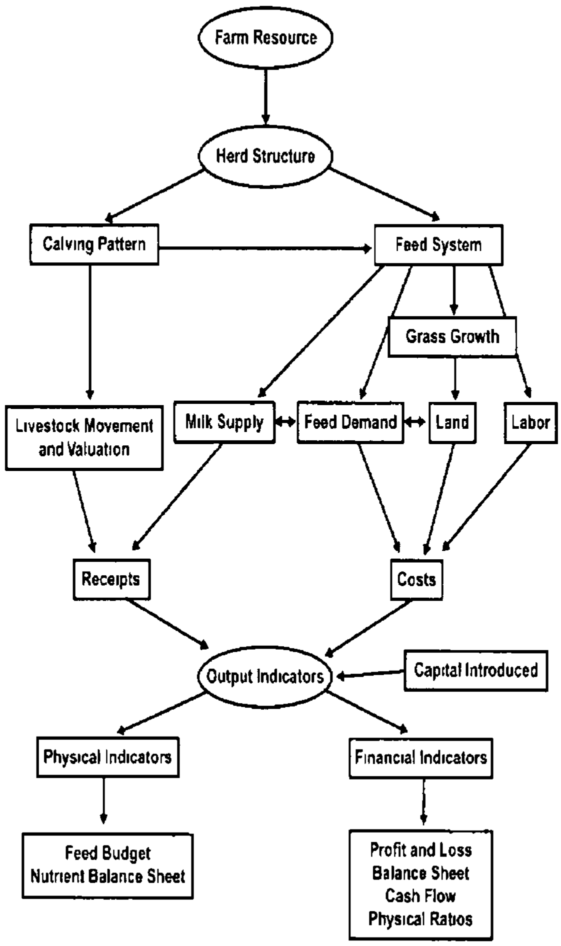


Figure 8 2 A flow chart of the Moorepark Dairy Systems Model

*Source* Shalloo et al (2004)

demand section of the model The proposed model will estimate the weight of a cow is at a certain time in lactation so that the optimum feed requirement can then be calculated Thus the costs and receipts of the farm as a whole can be calculated with greater precision than heretofore This is of great benefit to producers as they face important issues relating to improving efficiency, lowering costs and increasing productivity while being cognisant of issues related to the environment, animal welfare and food safety (Shalloo et al , 2004)

Furthermore, if one has a good understanding and an accurate model for predicting milk yield, fat content, protein content and liveweight in conjuction with the MDSM then one can

1 establish an accurate cost of milk production for each month of calving taking account of local environmental factors,

2 determine the optimum calving pattern and lactation length for spring and autumn calving herds taking cognisance of milk composition and processing qualities,

3 carry out further research on milk pricing systems that better reflect processing capabilities both in terms of milk supply pattern and processing qualities

The models provided in this thesis could be used to outline the optimum calving pattern, lactation profiles, milk composition and processing qualities for milk production systems

Currently, the national milk supply profile indicates that a peak to trough ratio of eight to one occurs in the monthly supply of milk to dairy processing plants, with the peak occurring in the months of May and June and the trough occurring in December and January (Central Statistics Office, 2005) (See Figure 8 3) The high peak to trough ratio results in processors requiring much more plant capacity than would be required if there was a lower peak Traditionally, an inconsistency in the processing quality of milk has existed in the trough months, which, given the

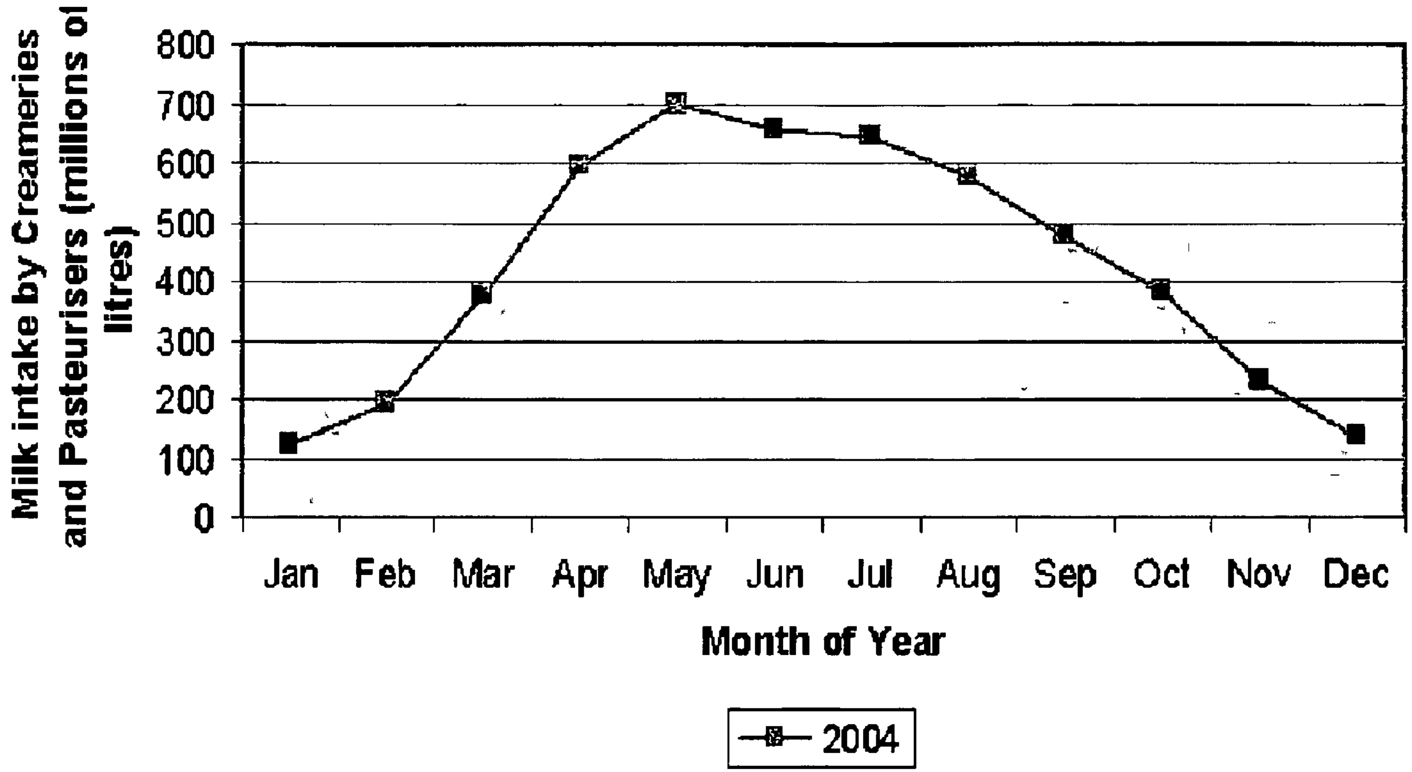


Figure 8 3 Summary of milk supply

*Source* Central Statistics Office (2005) calving pattern of the national herd, are supplied mainly from animals in early and late lactation

In early lactation the main problem is low protein concentration, while in late lactation the balance of fat, lactose, protein and casein content is altered Thus, seasonal variation in the composition of milk supplied affects the composition and quality of dairy products, resulting in the production of certain products becoming seasonal also Therefore the seasonality within the Irish dairy industry influences the mix of products, which can be manufactured from milk (O'Brien et al , 1996) The ability of processors to develop and produce high-value products all year round is reduced by a highly seasonal supply pattern

The composition and processability of milk in both early and late lactation can also be affected by the level and plane of nutrition (O'Brien et al , 1996, Crosse et al , 2000) Research has shown that consistency in milk composition and quality can be maintained up to 275 days in lactation within the Moorepark Blueprint for spring milk production (O'Brien, 1999) Within this system animals finish their lactation at a high level of milk production because they have adequate levels of nutrition As a result it is necessary to examine the influence of calving pattern, lactation length and nutritional level on milk composition as well as on processability In order to investigate these factors further it is necessary to have accurate models to predict the yield of milk and its constituents as they are the foundations on which accurate economic modelling of the Irish dairy industry is based

Milk pricing strategies generally reflect the level of fat and protein in milk as well as its microbiological quality (O'Brien et al , 1996) Currently a high price is paid for late lactation milk due to the high proportions of milk constituents, while the quality for processing is not considered The value of protein and fat in milk can be directly attributed to the sale of products from the processor It has been suggested that milk pricing needs to reflect more the processing characteristics of the milk (van Bekkum and Nilsson, 2000) In a study carried out by Breen (2001), milk price was calculated based on A+B-C, where A was protein value, B was fat value and C was the processing and collection costs The value of the protein and the fat was calculated monthly based on the product mix Breen (2001) developed a linear programming model the objective function of which was to maximize the milk price subject to the the constraints of plant capacity for individual products and milk supply profile Therefore the linear programming model maximizes the amount of higher-value products (such as functional foods and value added products) subject to the capacity of the plant to manufacture them If there was a more favorable supply profile it would be possible at processor level to produce more higher-value products and therefore the fat and protein value would be higher, resulting in a higher milk price

The models outlined in this thesis when amalgamated into the Moorepark Dairy Systems Model (MDSM) are of great benefit to the Irish dairy industry at a time when many challenges are facing it The Prospectus Report (Promar International, 2003) outlines three key strategies for the long term success of the Irish dairy industry (1) improvement of the international competitiveness, scale and cost efficiency of both the producers and processing sectors, (2) increase in the proportion of output away from base/commodity type products and into higher value-added products, and (3) greater emphasis being put on actions to develop and underpin the highest standards of quality and safety of Irish dairy produce Therefore as the Irish dairy industry is examining which route to take to secure the livelihoods of milk producers and processors, the next step is to use the models outlined in this study to model the optimum calving pattern, lactation profiles, milk composition and processing qualities for milk production systems under different economic conditions The price processors should pay producers to maintain a more even milk supply curve could then be estimated Thus, it is evident that the models outlined in this thesis are the foundation blocks on which very important decisions can be made at a crucial time in the Irish dairy industry

The models developed in this thesis offer a simple robust method for predicting milk yield, fat content, protein content and liveweight of Irish dairy cows. This has been done by using empirical regression techniques to model the data. Furthermore, it has been shown that an analysis of the residuals is important to examining regression models. It was evident that the model with the best MSPE value was not necessarily the most satisfactory model as there could be severe problems with the presence of multicollinearity among the variables. This was rectified by either removing a variable or by using PCA. Thus, the applicability of a model was based on its goodness-of-fit, its ability to adhere to the assumptions of regression analysis and its ability to predict either total milk yield or the average fat content, protein content or liveweight.

In closing, this thesis is an important step in modelling the lactation curves of cows by addressing the assumptions of regression analysis in detail, as well as the initial step in overcoming the major problems that milk processing plants have with the irregular supply of milk from farmers.

# BIBLIOGRAPHY


Agresti, A (1990) *Categorical Data Analysis* John Wiley & Sons, 2nd edition

Ali, T and Schaeffer, L (1987) Accounting for covariance among test day yields in dairy cows *Canadian Journal of Animal Science*, 67 637–644

Anderson, T (2002) Specification and misspecification in reduced rank regression *The Indian Journal of Statistics*, 64 193–205

Ayala, A , Honhold, N , Delgado, R , and Magana, J (1992) *A Visual Condition Scoring Scheme for Bos Idicus and Crossbred Cattle* Dual Purpose Cattle Production Research, Merdia, Mexico

Bakker, H and Koops, W (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 & Sons

Berglund, B and Danell, B (1987) Live weight changes, feed consumption, milk yield and energy balance in dairy cattle during the first period of lactation *Acta Agricultura Scandinavia*, 37 495–509

Berry, D , Horan, B , and Dillon, P (2005) Comparison of growth curves of three strains of female dairy cattle *Animal Science*, 80 151–160

Bowerman, B and O'Connell, R (1987) *Time Series Forecasting - Unified Concepts and Computer Implementations* Duxbury Press, Boston

Breen, J (2001) A new direction for the payment of milk Technology and seasonality considerations in multiple component milk pricing of milk (liquid and manufacturing) for a diversifying dairy industry Master's thesis, University College Dublin

Brody, S (1945) *Bioenergetics and Growth* Rheinhold Pub Corp , New York

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

Brotherstone, S , White, I , and Meyer, K (2000) Genetic modelling of dairy milk yield using orthogonal polynomials and parametric curves *Animal Science*, 70 407–415

Brown, J , Fitzhugh, H , and Cartwright, T (1976) A comparison of nonlinear models for describing weight-age relationships in cattle *Animal Science*, 42 810–818

Buckley, F , O'Sullivan, K , Mee, J , Evans, R , and Dillon, P (2003) Relationships among milk yield, body condition, cow weight and reproduction in spring-calved Holstein-Friesians *Journal of Dairy Science*, 86 2308–2319

Butler, S , Marr, A , Pelton, S , Radcliff, R , Lucy, M , and Butler, W (2003) Insulin restores GH responsiveness during lactation induced negative energy balance in dairy cattle effects on expression of IGF-I and GH receptor 1A *Journal of Endocrinology*, 176 205–217

Carvalheira, J , Blake, R , Pollak, E , and Quaas, R (1998) Application of an autoregressive process to estimate genetic parameters and breeding values for daily milk yield in a tropical herd of Lucerna cattle and in United States Holstein herds *Journal of Dairy Science*, 81 2738–2751

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

Central Statistics Office (2005) *Principal Statistics in Relation to Agriculture - Milk Statistics* Central Statistics Office, Cork

Cobby, J and Le Du, Y (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 , O'Brien, B , and Ryan, G (2000) Production of quality milk for processing *International Journal of Dairy Technology*, 54 135–141

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 and Food Research*, 27 31–38

Cunningham, E (1972) Components of variation on dairy cow production *Irish Journal of Agricultural and Food Research*, 11 1–10

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

De Boer, J , Weller, J , Gipson, T , and Grossman, M (1989) Multiphasic analysis of milk and fat yield curves of Israeli Holsteins *Journal of Dairy Science*, 72 2143–2152

De Vries, M and Veerkamp, R (2000) Energy balance of dairy cattle in relation to milk production variables and fertility *Journal of Dairy Science*, 82 1927–1934

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

Department of Agriculture and Food (2002) *Compendium of Agricultural Statistics 2002 - Update of 2001* Department of Agriculture and Food Research Development, Dublin, Ireland

Department of Agriculture and Food (2004) *Compendium of Agricultural Statistics 2004 - Update of 2003* Department of Agriculture and Food Research Development, Dublin, Ireland

Devir, S , Zur, B , Maltz, E , Genizi, A , and Antler, A (1995) A model for the prediction of dairy cow body weight based on a physiological timescale *Journal of Agricultural Science, Cambridge*, 125 415–424

Dhanoa, M (1981) A note on an alternative form of the lactation model *Animal Production*, 32 349–351

Draper, N and Smith, H (1981) *Applied Regression Analysis* John Wiley & Sons, 2nd edition

Elston, D , Glasbey, C , and Neilson, D (1989) Non-parametric lactation curves *Animal Production*, 48 331–339

Field, A (2000) *Discovering Statistics - Using SPSS for Windows* SAGE Publications

Fischer, A (1958) Research with Wurttemberg Spotted Mountain cows on the shape of the lactation curve and how it may be influenced by non-genetic factors *Zuchtungskunde*, 30 296–304

Forbes, J (1983) Models for the prediction of food intake and energy balance in dairy cows *Livestock Production Science*, 10 149–157

Fulkerson, W , Wilkins, J , Dobos, R , Hough, G , Goddard, M , and Davidson, T (2001) Reproductive performance in Holstein-Friesian cows in relation to genetic merit and level of feeding when grazing *Animal Science*, 73 397–406

Gaines, W (1927) Persistency of lactation in dairy cows *Bull Ill agric Exp Stn* , 288 355–424

Garcia, S and Holmes, C (2001) Lactation curves of autumn- and spring-calved cows in pasture-based dairy systems *Livestock Production Science*, 68 189–203

Gonzalo, C , Carriedo, J , Baro, J , and San Primitivo, F (1994) Factors influencing variation of test day milk yield, somatic cell count, fat and protein in dairy sheep *Journal of Dairy Science*, 77 1537–1542

Gravir, K (1967) Studies on different body measurements as estimators of live and carcass weight in young NRF bulls *Acta Agricultura Scandinavia*, 17 217–227

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 (1995) Modelling of lactation curve as a sub-model in the evaluation of test day records In *Proceedings of INTERBULL open meeting*, Prague, Czech Republic

Guthrie, L (1994) *Factors that Influence Milk Fat Tests* The University of Georgia College of Agricultural and Environmental Sciences, 836 edition

Heinrichs, A , Rogers, G , and Cooper, J (1992) Predicting body weight and wither height in Holstein heifers using body measurements *Journal of Dairy Science*, 75 3576–3581

Heuer, C , Schukken, Y , and Dobbelaar, P (1999) Post partum body condition score and results from the first test day milk as predictors of disease, fertility, yield and culling in commercial dairy herds *Journal of Dairy Science*, 82 295–304

Huggett, A and Widdas, W (1951) The relationship between mammalian foetal weight and conception age *Journal of Physiology*, 114 306–317

ICAR (2002) *Recording Guidelines International Agreement of Recording Practices* International Committee of Animal Recording

ICBF (2003) *Irish Cattle Breeding Statistics 2002* Irish Cattle Breeding Federation, Highfield House, Shinagh, Bandon, Co Cork, Ireland

ICBF (2004) *ICBF's Electronic DIY Milk Recording System* Irish Cattle Breeding Federation, Highfield House, Shmagh, Bandon, Co Cork, Ireland

ICBF (2005) Personal communication

Jamrozik, J , Gianola, 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

Jamrozik, J , Kistemaker, G , Dekkers, J , and Schaeffer, L (1997) Comparison of possible covariates for use m a random regression model for analyses of test day yields *Journal of Dairy Science*, 80 2550–2556

Jamrozik, J and Schaeffer, L (1997) Estimates of genetic parameters for a test day model with random regressions for yield traits of first lactation Holsteins *Journal of Dairy Science*, 80 762–770

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

Kavanagh, S , Maher, J , Shalloo, L , and Kelly, F (2003) Cost effectıve feedıng systems for daıry cows In *Proceedıngs of the Natıonal Daıry Conference 2003*, Rochestown Park Hotel, Cork

Kellogg, D , Urquhart, N , and Ortega, A (1977) Estımatıng Holsteın lactatıon curves wıth a gamma functıon *Journal of Daıry Scıence*, 60 1308–1315

Keown, J and Van Vleck, L (1973) Extended lactatıon records ın progress to 305-day equıvalent *Journal of Daıry Scıence*, 56 1070–1079

Kettunen, A , Mantysaarı, E , and Poso, J (2000) Estımatıon of genetıc parameters for daıly mılk yıeld of prımıparous Ayshıre cows by random regressıon test-day models *Lıvestock Productıon Scıence*, 66 251–261

Kıllen, L and Keane, M (1978) The shape of lactatıon curves ın Irısh daıry herds *Irısh Journal of Agrıcultural and Food Research*, 17 267–282

Klopcıc, M , Malovrh, S , Gorȷanc, G , Kovac, M , and Osterc, J (2003) Predıctıon of daıly mılk, fat and proteın content usıng alternatıng recordıng scheme *Czech Journal of Anımal Scıence*, 48 449–458

Koonawootrıttrıron, S , Elzo, M , Tumwasorn, S , and Sıntala, W (2001) Lactatıon curves and predıctıon of daıly and accumulated mılk yıelds ın a multıbreed daıry herd ın Thaıland usıng all daıly records *Thaı Journal of Agrıcultural Scıence*, 34 123–139

Korver, S , Van Arendonk, J , and Koops, W (1985) A functıon for lıve-weıght change between two calvıngs ın daıry cattle *Anımal Productıon*, 40 233–241

Kumar, R and Bhat, P (1979) Lactatıon curve ın Indıan buffaloes *Indıan Journal of Daıry Scıence*, 32 156–160

Kvanlı, A , Guynes, C , and Pavur, R (1986) *Introductıon to Busıness Statıstıcs* West Publıshıng Company

LeBlanc, D (2004) *Statistics - Concepts and Applications for Science* Jones and Bartlett Publishers

Lee, C , Pollak, J , and Everett, R (1995) Multiplicative factors for estimation of daily milk and component yields from single morning or afternoon tests *Journal of Dairy Science*, 78 221–235

Lidauer, M , Mantysaari, E , Straden, I , and Poso, I (2000) Multiple-trait random regression test-day model for all lactations In *Proceedings of INTERBULL meeting*, pages 81–86

Lopez-Villalobos, N , Lemus-Ramirez, V , Holmes, C , and Garrick, D (2001) Lactation curves for milk traits, live weight and body condition score for heavy and light Holstein-Friesian cows In *Proceedings of New Zealand Society of Animal Science*, volume 61, pages 217–220

Macciotta, N , Vicario, D , and Pulma, G (2002) Test day and lactation yield predictions in Italian simmental cows by ARMA methods *Journal of Dairy Science*, 85 3107–3114

Mackle, T , Bryant, A , Petch, S , Hooper, R , and Auldist, M (1999) Variation in the composition of milk protein from pasture-fed dairy cows in late lactation and the effect of grain and silage *New Zealand Journal of Agricultural Research*, 42 147–154

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

Madalena, F , Teodoro, R , and Madureira, A (2003) Relationships of weight and height with age in hybrid Holstein Friesian /Guzera females *Genetic Molecular Research*, 2 271–278

Maddala, G (1992) *Introduction to Econometrics* MacMillan Publishing Company, New York

Maloney, J (2002) Higher milk prices and value added foods In *Proceedings of the National Dairy Conference 2002*, Gleneagle Hotel, Killarney, Ireland

McCarthy, D (2000) Optimising profits on dairy farms In *Proceedings of the National Dairy Conference 2000*, Silver Springs Hotel, Tivoli, Cork, Ireland

McGuire, M , Theurer, M , Vicini, J , and Crooker, B (2004) Controlling energy balance in early lactation *Advances in Dairy Technology*, 16 241–252

Mendoza, G (1999) *Serial Autocorrelation Why it's here, what it does and how to get rid of it* Fordham University, New York

Molina, J and Boschini, C (1979) Adjustment of dairy curve of Holstein herd with a linear modal model *Agronomia Costarricense*, 3 167–174

Moore, A (1985) A mathematical equation for animal growth from embryo to adult *Animal Production*, 40 441–453

Mukundan, G and Bhat, P (1983) Lactation curve in Malabari goats and their Saanen half-breds *Indian Journal of Animal Science*, 53 666–669

O'Brien, B (1999) *Chemical Composition and Processability of Milks from Herds with Different Calving Patterns - Project number 4349* Teagasc

O'Brien, B , Crosse, S , and Dillon, P (1996) Effects of offering a concentrate or silage supplement to grazing dairy cows in late lactation on animal performance of spring calving dairy cows *Irish Journal of Agricultural and Food Research*, 35 113–125

Olori, V , Brotherstone, S , Hill, W , and McGuirk, B (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 and Galesloot, P (1999) Projection of partial lactation records and calculation of 305-day yields in the Republic of Ireland In *Proceedings of the 1999 INTERBULL meeting*, pages 149–154, Zurich, Switzerland

Papajcsik, I and Bodero, J (1988) Modelling lactation curves of Friesian cows in a subtropical climate *Animal Production*, 47 201–207

Perochon, L , Coulon, J , and Lescourret, F (1996) Modelling lactation curves of dairy cows with emphasis on individual variability *Animal Science*, 63 189–200

Perotto, D , Cue, R , and Lee, A (1992) Comparison of nonlinear functions for describing the growth curve of three genotypes of dairy cattle *Canadian Journal of Animal Science*, 72 773–782

Promar International (2003) *Strategic Development Plan for the Irish Dairy Processing Sector* Promar International

Ptak, E and Schaeffer, L (1993) Use of test day yields for genetic evaluation of dairy sires and cows *Livestock Production Science*, 34 23–34

Pulina, G (1990) L'influenza dell'alimentazione sulla qualita del latte ovino (influence of feeding on sheep quality milk) *L'informatore Agrario*, 46 31–38

Quinn, N , Killen, L , and Buckley, F (2004) Statistical modelling of lactation curve data In *Proceedings of CompStat 2004*, pages 1677–1684, Prague, Czech Republic

Quinn, N , Killen, L , and Buckley, F (2005a) Empirical algebraic modelling of lactation curves using Irish data *Irish Journal of Agricultural and Food Research*, 44 [In Press]

Quinn, N , Killen, L , and Buckley, F (2005b) Modelling lactation curves for Irish dairy cows In *Performance Recording of Animals - State of the Art 2004*, pages 247–252, Sousse, Tunisia

Reents, R , Dekkers, J , and Schaeffer, L (1995) Genetic evaluation for somatic cell score with a test day model for multiple lactations *Journal of Dairy Science*, 78 2858–2870

Reents, R , Dopp, L , Schmutz, M , and Reinhardt, M (1998) Impact of application of a test day model to dairy production traits on genetic evaluations of cows In *Proceedings of the INTERBULL Meeting*, pages 49–54, Uppsala, Sweden

Rekaya, R , Carabano, M , and Toro, M (2000) Bayesian analysis of lactation curves of Holstein-Friesian cattle *Journal of Dairy Science*, 83 2691–2701

Reksen, O , Havrevoll, O , Grohn, Y , Bolstad, T , Waldmann, A , and Ropstad, E (2002) Relationships among body condition score, milk constituents, and postpartum luteal funstion in Norwegian dairy cows *Journal of Dairy Science*, 85 1406–1415

Richards, F (1959) A flexible growth function for empirical use *Journal of Experimental Botany*, 10 290–300

Roche, J (2003) Effect of pregnancy on milk production and body weight from identical twin study *Journal of Dairy Science*, 86 777–783

Rook, A , France, J , and Dhanoa, M (1993) On the mathematical description of lactation curves *Journal of Agricultural Science*, 121 97–102

Rowlands, G , Lucey, S , and Russel, A (1982) A comparison of different models of the lactation curve in dairy cattle *Animal Production*, 35 135–144

SAS (1999) *SAS Online Version 8* SAS Institute, Inc Cary NC

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

Schaeffer, L and Dekkers, J (1994) Random regressions in animal models for test-day production in dairy cattle In *Proceedings of 5th World Congress on Applied Livestock Production*, volume 18, pages 443–446, Guelph, Canada

Schaeffer, L and Jamrozik, J (1996) Multiple-trait prediction of lactation yields for dairy cows *Journal of Dairy Science*, 79 2044–2055

Schaeffer, L , Jamrozik, J , Kistemaker, G , and Van Doormaal, B (2000) Experience with test-day model *Journal of Dairy Science*, 83 1135–1144

Schutz, M , Hansen, L , Steuernagel, G , and Kuck, A (1990) Variation in milk, fat, protein and somatic cells for dairy cattle *Journal of Dairy Science*, 73 484

Sen, A and Srivastava, M (1990) *Regression Analysis - Theory, Methods and Applications* Springer-Verlag

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*, 87 1945–1959

Shanks, R , Berger, P , Freeman, A , and Dickenson, F (1981) Genetic aspects of lactation curves *Journal of Dairy Science*, 64 1852–1860

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 (1950) A study of lactation as affected by heredity and environment *Journal of Dairy Science*, 17 231–252

Singh, B and Bhat, P (1978) Models of lactation curves for Hariana cattle *Indian Journal of Animal Sciences*, 48 791–797

Singh, R and Gopal, R (1982) A study of lactation as affected by heredity and environment *Journal of Dairy Research*, 17 231–252

Slater, K and Webster, J (2001a) *Strategies to Modify Milk Fat* DEFRA

Slater, K and Webster, J (2001b) *Strategies to Modify Milk Protein* DEFRA

Stanton, T , Jones, L , Everett, R , and Kachman, S (1992) Estimating milk, fat and protein lactation curves with a test day model *Journal of Dairy Science*, 75 1691–1700

Taylor, S (1980) Genetically standardized growth equations *Animal Production*, 30 167–175

Tozer, P and Huffaker, R (1999) Mathematical equations to describe lactation curves for Holstein-Friesian cows in New South Wales *Australian Journal of Agricultural Research*, 50 431–440

van Bekkum, O and Nilsson, J (2000) Agenda 2000 and structural change in european dairy cooperatives In *Proceedings of the IVth International Conference on Chain Development in Agribusiness, and the Food Industry*, Wagemngen, Netherlands

Vargas, B , Koops, W , Herrero, M , and Van Arendonk, J (2000) Modeling extended lactations of dairy cows *Journal of Dairy Science*, 83 1371–1380

Varona, L , Moreno, C , Garcia Cortes, L , and Altarriba, J (1998) Bayesian analysis of Wood's lactation curve for Spanish dairy cows *Journal of Dairy Science*, 81 1469–1478

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 an autoregressive test day model In *Proceedings of the 7th World Congress on Genetics Applied to Livestock Production*, pages 279–282, Montpellier, France

Vollebregt, L and Vollebregt, R (1998) *Low Replacement Rates for Profit*

Von Bertalanffy, L (1957) Quantitative laws in metabolism and growth *Quarterly Review of Biology*, 32 217–231

Walter, J , Mao, I , and Emery, R (1984) Simultaneous fitting of energy requirements and intake, lactation response and body weight changes in dairy cattle with different forage basses rations *Journal of Dairy Science*, 70 221

Wicks, H (2001) *The Influence of Genetic Merit and Environment on Dairy Cattle Performance on Farm* PhD thesis, Imperial College of Wye, University of London

Wiggans, G , VanRaden, P , and Philpot, J (2003) Technical note Detection and adjustment of abnormal test-day yields *Journal of Dairy Science*, 86 2721–2724

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

Winsor, C (1932) The Gompertz curve as a growth curve In *Proceedings of the National Academy of Sciences*, volume 18, pages 1–8, USA

Wood, P (1967) Algebraic model of the lactation curve in cattle *Nature, London*, 216 164–165

Wood, P (1969) Factors affecting the shape of the lactation curve in cattle *Animal Production*, 11 307–316

Wood, P (1976) Algebraic models of the lactation curves for milk, fat and protein production, with estimates of seasonal variation *Animal Production*, 22 35–40

Wood, P , King, J , and Youdan, P (1980) Relationships between size, liveweight change and milk production characters in early lactation in dairy cattle *Animal Production*, 31 143–151

Yadav, M , Katpatal, B , and Kaushik, S (1977) Components of inverse polynomial function of lactation curve, and factors affecting then in Harina and its Friesian crosses *Indian Journal of Animal Science*, 47 777–781