Environmental adjustments in estimation of breeding value of dairy sires

by

Vijay P.S. Chauhan

Thesis presented for the degree of
Doctor of Philosophy
University of Edinburgh
1985

Declaration

I declare that the research work included in this thesis is my own and no parts of it have been presented for any previous degree.

Acknowledgements

Environmental adjustments in estimation of breeding value of dairy sires

INDEX

ABSTRACT

Alternative models for the adjustment of environmental effects in dairy sire evaluation were investigated in order to overcome some of the limitations of the commonly used herd-year-season fixed effects model.

The data comprised the first lactation records on 49242 progeny of 69 widely used proven Holstein-Friesian sires in England and Wales. The effects of sire (as fixed, since the data were on proven sires), overall effects of year and month of calving (both as fixed) and age as a covariable by linear and quadratic regressions were the same in all models. Altogether 5 models which differed in fitting the following effects were examined: (1) herd-year-season fixed; (2) herd-year fixed and herd-year-season random; (3) herd fixed, herd-year and herd-year-season random; (4) herd-year fixed, with the variance-covariance structure between months included; (5) herd-year fixed and herd-year-month-age class random. Seasons were taken to be of either 1 or 2 or 4 months.

Using model 1 the variances of the estimates of sire effects were considerably larger when seasons were shorter, showing that shorter seasons (fixed) result in substantial losses of information. Longer seasons introduced larger herd-year X month within season interactions. Models 2 and 3 were observed to be more efficient than model 1. They increased the accuracy, not just from the recovery of inter-season information but also from the use of grouping of one month's duration. As the seasons (random) were assumed to be shorter the advantages of using random effects models became larger. Model 4 did not show much improvement over model 2 with each season of one month. This model was found to give larger advantages when herds were bigger, however, it was observed to be only 2.6% more efficient than model 2 with the maximum average herd-year size (14.8) in the data sets created for large herd situations. Model 5 was chosen in order to account for any biases due to interactions of age at calving with herd, year or season, however the variance component for age within herd-year-month was not different from zero.

It was concluded that a model similar to either model 2 or 3 will be useful for sire evaluation, particularly in small herds such as those in Switzerland and India. These two models each using seasons (random) of one month overcome the limitations of seasonal groupings.

Chapter 1

GENERAL INTRODUCTION

Selection of sires through progeny test has been the main basis for genetic improvement of dairy cattle. Only a maximum of 10% of the variation in milk, fat and protein yields is accounted for by sire effects, however, therefore much of the variation observed in milk production is non-genetic or environmental. As is evident from the literature (Chapter 2), in the estimation of the breeding values of dairy sires little attention has traditionally been paid to the analyses of the environmental factors causing variation in milk production, therefore the main objective of this group of studies is to investigate alternative models for the adjustment of environmental effects in a dairy sire evaluation model. The effects of herd, year, season, age of calving and interactions among them are the identifiable factors which account for most of the environmental variation in milk production. In order to account for the interactions between the effects of herd, year, and season these effects are fitted as a joint effect (i.e. the herd-year-season) in the model for sire evaluation used in most breeding schemes, and data are precorrected for the effects of age and month of calving using standard correction factors derived from large data sets. Seasons are usually assumed to be 2 or 3 arbitrary groups of equal or unequal number of consecutive months and herd-year-season effects are fitted as fixed.

The herd-year-season fixed effects model is widely used and it has become a conventional model for estimating the breeding values of dairy sires. The main reasons that can be given for the above practices are (i) herd-year-season effects are assumed to be fixed in order to remove any possible bias that could arise from the non-random use of bulls across herds, (ii) milk yield varies between months, but months have to be grouped into seasons to avoid losses of information due to small environmental subclasses in herd-year-month models, (iii) precorrections for age and month of calving are done to reduce the amount of computation.

The above procedure of environmental adjustments ignores the herd-year-season x month and herd-year-season X age of calving interactions. Because herd-year-seasons are fitted as fixed effects, the covariances between cows in the same or different seasons in the same herd are assumed to be zero. Although it is reasonable to assume the herd effects to be fixed for the non-random use of bulls across herds because some farmers tend to have preferences for particular bulls, there do not seem to be any explanations why the year effects within herd and season effects within herd-year are assumed to be fixed. However, if the year trends within herd are important then it may be useful to assume herd-year effects to be fixed, but the season effects within herd-year (i.e. the herd-year x season interaction) can still be assumed to be random. Since the seasonal and management variations from month to month are generally large, long seasonal groupings do not seem to be appropriate for progeny group comparisons. The cows calving in the beginning of a long season and those calving towards the end hardly perform under similar environments. This procedure

2

of splitting cows by their calving dates into different herd-year-season subclasses or "contemporary groups" is such that a few hours difference in the calving times would send them into different contemporary groups. Therefore a model with shorter seasons may be more suitable than the models assuming seasons of 4-6 months duration.

Whatever the length of seasons may be, the fixed season models may not be efficient because the covariances between seasons are assumed to be zero. If each season comprised one month then the covariance of cows calving in a given month with the cows calving in the preceding and the following months may be higher than with those calving more months apart. Therefore, a model in which the environmental covariances between months are accounted for would also be desirable.

The objectives of this work are to examine some of the above limitations of the herd-year-season fixed effects model of sire evaluation, and to suggest some alternative models for adjustment of environmental effects. Specifically, the following areas are chosen:

(1) to examine the criteria of seasonal grouping in a herd-year-season fixed effect model of sire evaluation which would minimise herd-year-season X month interactions, so that precorrections for month effects in all herds are efficient (Chapter 3),

(2) to estimate environmental correlations between records within seasons of variable lengths, in order to obtain the variance

components necessary for random effects analyses (Chapter 4),

(3) to contrast the herd-year-season fixed effects model with models in which herd is fixed, and the effects of year within herd and season (different lengths) within herd-year are assumed to be random (Chapter 5),

(4) to estimate the environmental covariances between months within herd-year, and to fit a sire evaluation model that accounts for these covariances (Chapters 6 and 7),

(5) finally, to suggest a model for sire evaluation suitable in very small herd situations, for example, India (Chapter 8).

The findings from the literature on the above areas will be discussed in Chapter 2. However, parts of the review will also be included in the subsequent chapters where appropriate.

Chapter 2

LITERATURE REVIEW

The non-genetic parts of the variance in milk production comprise the environmental and residual variances. The identifiable environmental factors causing variation are herd, year, month of calving, age at calving, lactation length, milking frequency and interactions among them, etc. When an environmental effect is ignored it may result either in an increased residual variance and/or in partial confounding of its variation with other effects. Consequently, the sire evaluations may be biased.

VARIATION DUE TO DIFFERENT ENVIRONMENTAL EFFECTS:

The variance due to herd, year-season, sire and interactions among them reported from several studies (Hickman and Henderson, 1955; Legates, Verlinden and Kendrick, 1956; Van Vleck, Wadell and Henderson, 1961; Allaire and Gaunt, 1965; Bereskin and Freeman, 1965; Harville and Henderson, 1967; Fimland, Bar-Anan and Harvey, 1972b) are given in Table 2.1. Most of the reports given in this table are from the United States in which the herd components were observed to be about 30% for milk and fat yields and about 15% for fat contents. However, the proportions of variance due to herd reported for Friesians in Great Britain (Robertson and Rendel, 1954; Robertson and Khishin, 1958; and Barker and Robertson, 1966) and for Israeli Friesians (Fimland et al, 1972b) are slightly smaller. Unfortunately, the results of different studies are not directly

comparable since the models of analyses and the bases of choosing data sets often differ.

The herd component of variance consists of both genetic and environmental portions. The genetic portion is, however, small (about 10% of the total herd component) as reported by Lush and Straus (1942), Robertson and Rendel (1954), Robertson and McArthur (1955), Freeman and Henderson (1959), Pirchner and Lush (1959) and Moriello and Legates (1970). The genetic variation among herds is mainly due to use of natural service sires and non-random use of A.I. sires, i.e. superior sires being used in high yielding herds. Because of extensive use of artificial insemination and advanced selection practices in dairy cattle, the use of natural service sires is limited , at present. A model with an effect of maternal grandsire (Quass, Everett and McClintock, 1979) is used in some breeding schemes (e.g. the MMB) in order to account for genetic variation between the dams of progeny.

Van Vleck et al (1961) and Bereskin and Freeman (1965) reported that the variance component for year-season was about 2% of the total variance in milk and fat yields while the estimates reported by Hickman and Henderson (1955) and Harville and Henderson (1967) were slightly higher (about 5%). In a study ignoring year X season interactions, Gacula, Gaunt and Damon (1968) reported that the year components for yield and composition traits were respectively 1.0 and 1.9%, and season components 1.9 and 1.4% of the total variance. The effects of month of calving on milk, fat and protein yields and fat and protein contents have been reported to be significant (Sanders, 1927; Bereskin and Freeman, 1965; Sargent, Butcher and Legates, 1967; Wood, 1970; Norman, Kuck, Cassell and Dickinson,

6

1978; Cooper and Hargrove, 1982; McClintock, 1982). Bereskin and Freeman (1965) reported the month component to account for 1.4% of the total variance in milk and 1.7% in fat yields.

In contrast to the year-season effects across herds the year-season effects within herd (i.e. the herd X year-season interaction) typically account for three times as much variation as the year-season effects alone (Hickman and Henderson, 1955; Van Vleck et al, 1961). The variance due to year effects within herds reported by Barker and Robertson (1966) was similar to the combined effects of year-season and herd x year-season interactions reported by Van Vleck et al, (1961) and Harville and Henderson (1967).

Age effects have been found to be highly significant in many studies (Fimland, Bar-Anan and Harvey, 1972a; Norman et al, 1978; Cooper and Hargrove, 1982). The age component has been reported to account for about 20 - 25% of the total variance in milk and fat yields and 2 - 5% in fat and protein contents, respectively (Robertson, Waite and White, 1956; Gacula et al, 1968; Sargent et al, 1967). Age X season of calving interactions were also reported to be significant in several studies (Van Vleck and Henderson, 1961; Syrstad, 1965; Wunder and McGilliard, 1967; Mao, Burnside, Wilton and Freeman, 1974; and Norman et al, 1978). According to these authors, and also Miller, Lentz and Henderson (1970) and Lee and Hickman (1972), joint corrections for age and month of calving would be desirable.

Fimland et al (1972a) reported that the herd-year-season X age interactions were significant for milk and fat yields. The herd x age interaction is mainly due to differences in the maturity of the

cows between herds, in other words all breeders do not breed their heifers at the same age. Those who raise their heifers on a good plane of nutrition would tend to breed them at an early age. In an attempt to partially overcome the problems of herd X age interactions, Searle and Henderson (1959) and Searle (1962) recommended the use of separate sets of age correction factors derived for different herd production levels. However, the component of variance for herd-year-season x age interaction was observed to be small (Fimland et al, 1972a). Barker and Robertson (1966) reported that the residual mean squares within herd-year-month-age subclasses were about two-thirds of those within the herd-year-month subclasses, therefore a herd-year-month-age model may be more suitable for sire evaluation.

ENVIRONMENTAL CORRELATIONS:

The covariances between records within different environmental subclasses are often expressed in terms of the environmental correlations. The significance of these correlations in sire evaluations depend on the assumptions made in the model. The herd-year-season effects are invariably assumed to be fixed in the sire evaluation model used in most breeding schemes. The herd effects are usually assumed to be fixed in order to account for the bias due to non-random use of sires across herds. However, if some of the environmental effects, e.g. season within herd-year or both season and year within herd, can be assumed to be random then the corresponding correlations among records in the herd-year and herd-year-season subclasses may be of interest from the point of view of evaluating the commonly used herd-year-season fixed effects model of sire evaluation.

Bereskin and Lush (1965) demonstrated the effects of environmental correlations, caused by factors other than sire, on the accuracy of prediction of breeding value (in the herd-mate comparison method). These authors indicated that the values of the accuracy were 0.91, 0.85, 0.81, 0.74 and 0.68 respectively when the environmental correlations were of the order of 0, 0.01, 0.02, 0.04 and 0.06. Therefore the accuracy of sire evaluation decreases as the environmental correlations increase.

Miller (1964), Thomson and Freeman (1970) and Arora and Freeman (1971) used the following model and the formulae for estimating correlations among records within different environmental subclasses. Each record was represented by the following model:

$$Y_{ijklm} = \mu + a_i + h_j + z_k + d_{jkl} + e_{ijklm}$$

where: μ is the overall mean and Y_{ijklm} is the mth record of the lth daughter (d) of the kth sire (z) made in the jth herd (h) and started in the ith year-season (a). All effects, fitted in the above described model and the residual error (e) were assumed to be random distributed with mean zero and the variances σ_a^2, σ_h^2, σ_z^2, σ_d^2 and σ_e^2, respectively. The formulae used for estimating correlations among records in the same herd and the same year-season (r_{ha}), in the same herd but different year-season (r_h), and in different herds but the same year-season (r_a) are given below:

$$r_{ha} \quad : \quad (\sigma_a{}^2 + \sigma_h{}^2)/(\sigma_a{}^2 + \sigma_h{}^2 + \sigma_z{}^2 + \sigma_d{}^2 + \sigma_e{}^2)$$

$$r_h \quad : \quad \sigma_h{}^2/(\sigma_a{}^2 + \sigma_h{}^2 + \sigma_z{}^2 + \sigma_d{}^2 + \sigma_e{}^2)$$

$$r_a \quad : \quad \sigma_a{}^2/(\sigma_a{}^2 + \sigma_h{}^2 + \sigma_z{}^2 + \sigma_d{}^2 + \sigma_e{}^2)$$

In their model they ignored the herd X year-season interaction effects (i.e. aXh interaction). This interaction effect should have been included in their model, however, for estimating the correlation among records within the same herd and the same year-season. Then the following formula would have been appropriate for estimating r^*_{ha},

$$r^*_{ha} = (\sigma_a{}^2 + \sigma_h{}^2 + \sigma_{aXh}{}^2)/(\sigma_a{}^2 + \sigma_h{}^2 + \sigma_{aXh}{}^2 + \sigma_z{}^2 + \sigma_d{}^2 + \sigma_e{}^2)$$

The environmental correlations reported by Thomson and Freeman (1970) and Arora and Freeman (1971) using mature equivalent milk and fat yield records are given below.

	r_{ha}	r_h	r_a
Thomson and Freeman (1970)			
milk yield:	0.328	0.226	0.028
Arora and Freeman (1971)			
milk yield:	0.357	0.198	0.022
fat yield:	0.381	0.216	0.018

If the variance components given in Table 2.1 were to be interpreted in terms of the environmental correlations, then the environmental correlations corresponding to the above notations are

as follows:

	r^{*}_{ha}	r_h	r_a
Hickman and Henderson (1955)			
milk yield:	0.483	0.300	0.040
fat field:	0.524	0.329	0.047
Van Vleck et al (1961)			
milk yield:	0.381	0.291	0.021
fat yield:	0.428	0.317	0.026
Harville and Henderson (1967)			
milk yield:	0.377	0.266	0.051
fat yield:	0.403	0.289	0.047

These correlations are larger than those of Thomson and Freeman (1970) and Arora and Freeman (1971). The estimates reported by these authors were biased downward presumably due to the fact that they ignored the herd X year-season interaction effects in their models and also due to culling as they used records of all lactations.

ENVIRONMENTAL GROUPING FOR PROGENY GROUP COMPARISONS

Environmental grouping in a sire evaluation model is a highly debatable issue. The environmental effects are grouped together arbitrarily in most breeding schemes. An efficient grouping is very important in order to make the best use of field data.

Grouping months to form seasons: Seasonal grouping is useful in

sire evaluation to increase the number of comparisons. If the environmental subclasses are small (e.g. herd-year-month) then there may be substantial losses of information.

Two types of seasonal groupings - fixed and moving seasons - are generally used in sire evaluation. Fixed seasons are used in almost all herd-year-season models of sire evaluation. A moving season over 5-months centered on the month of calving is also used by the USDA-DHIA in their Modified Contemporary Comparison method of sire evaluation (Dickinson, Norman, Powell, Waite and McDaniel, 1976). Fixed seasons are taken to be arbitrary groups of consecutive calendar months, in most breeding schemes. For example, the Northeast Artificial Insemination Sire Comparison (NEAISC) of New York uses two seasons: December-April and May-November (Bolgiano, Van Vleck and Everett, 1979), and the Milk Marketing Board of England and Wales (MMB) uses three seasons: December-March, April-July and August-November (G.J.T. Swanson, personal communication, 1984). In several studies it has been recommended that consecutive months with similar effects on milk production should be grouped together (Tucker and Legates, 1962; Gaunt, Bartlett and Comstock, 1964; Sargent, et al, 1967). Bereskin and Freeman (1965) suggested the criteria of largest between seasons, smallest within season and smallest residual variances, which are equivalent to grouping months with similar effects. However, there are no reports showing how the residual variance changes by grouping months with different trends in the milk production.

Fixed seasons, however, do not seem to be appropriate as the covariances among cows calving in the same and different months within season are assumed to be equal (given that the season effect

12

is assumed to be random). This limitation can be overcome to some extent by using shorter seasons or by using a model assuming covariances between months. Some models similar to those used in the comparisons of crop varieties in agricultural experiments (Bartlett, 1978; Wilkinson, Eckert, Hancock and Mayo, 1983) may also be desirable. Compared to a fixed seasons grouping in the herd-year-season model, smaller residual variances have been reported from analyses of records deviated from a moving season mean (Gaunt, et al, 1964; Bereskin and Freeman, 1965). The authors of the latter study reported that the residual variances for milk and fat% were smaller by factors of 0.90 and 0.57, respectively, but for fat yield the residual variance was larger by a factor of 1.28. Allaire and Gaunt (1965) also reported that the residual variances were smaller in records deviated from moving season means.

The moving seasons were presumably used in order to provide a similar variance-covariance structure between months for records of each month. Van Vleck (1966) reported that the correlations, estimated using records deviated from herd mean, between milk yields of unrelated Holstein cows calving 0 to 18 year-seasons apart did not show any trend as the distance between calving dates increased. Environmental correlations among records in the same year-season (i.e. 0-month-apart) and 1-, 2-, 5-, 10, 15- and 17-year-seasons apart were estimated to be 0.007, 0.081, 0.055, 0.073, 0.004, 0.090 and 0.099, respectively. Van Vleck himself considered these correlations to be unrealistic since he expected them to show a declining trend as the distance between year-seasons increased. These estimates are biased as they involve the covariances of each record with the rest in a herd, since each record was expressed as a deviation from the herd mean.

Grouping several environmental effects: While comparing different methods of sire evaluation, Dempfle (1977) discussed the uses of different environmental groupings under different herd size situations. He suggested that a herd-class effect (i.e. the grouping of herds according to milk production levels) could be more useful than the herd effect for small herds. It was also pointed out that the herd-year effects model may also be more efficient than the herd-year-season model for small herds. However, these suggestions were not supported by any results from the field data analyses. Considering the small herd problems in Switzerland, Hagger and Dempfle (1983) have suggested a region-herdclass-year-season (fixed) grouping rather than the usual herd-year-season (fixed) grouping. Although the model with region-herdclass-year-season effects gave a slightly larger residual variance than the herd-year-season model, it increased the number of comparisons substantially. It could be argued, however, whether it was necessary to have an environmental grouping fitted essentially as a fixed effect. A model in which some environmental effects are assumed as fixed and some effects as random would have been presumably more efficient than the region-herdclass-year-season fixed effects model. With this grouping they had to ignore the herd effects.

ADJUSTMENTS FOR THE ENVIRONMENTAL EFFECTS:

Procedures for adjustment of environmental effects used in different methods of estimating breeding values of sires are discussed below.

14

In several studies pertaining to the herd-mate comparison method of sire evaluation (Henderson, Carter and Godfrey, 1954; Henderson, 1956) the adjustments for environmental effects were observed to be more efficient when the covariances among records of the herd-mates were accounted for (Heidhues, Van Vleck and Henderson, 1961; Van Vleck, Heidhues and Henderson, 1961).

At present, the best linear unbiased predictor (BLUP) procedure is the method of choice for sire evaluation (Henderson, 1973, 1975b; Thompson, 1979). A model with herd-year-season as a fixed effect, sire effect as random and the group of sire as a fixed effect, is commonly used. In this model all covariances among records in the same or different seasons in the same herd are assumed to be zero, since herd-year-yeason is fitted as a fixed effect. Because of the non-random use of bulls across herds it seems reasonable to assume herd effects to be fixed. However, there is no evidence suggesting that the year effects within herd and/or the season effects within herd-year should not be assumed to be random. If the year and season effects within herd can be assumed to be random, then extra information on progeny group comparisons could be recovered from the inter-year and inter-season comparisons (Cunningham and Henderson, 1966; Patterson and Thompson, 1971).

Cunningham (1965) and Henderson (1973, 1975a, 1975b), in simple examples, and Miller, McDaniel and Plowman (1968) using field data, have compared sire evaluation models in which a single environmental factor was fitted as a fixed or random effect. They showed that a random effects model gave smaller sampling errors of the estimates of sire effects. Miller et al (1968) compared the least squares and maximum liklihood methods of sire evaluation. In their models they

fitted the effects of herd-year, season of calving, year of birth, season X year interaction, group of sire and sire within group. In the least squares method all effects except error (error: random with mean zero and variance σ_e^2) were assumed to be fixed. The maximum liklihood method differed from the least squares method in that the herd-year effects were assumed to be random distributed with mean zero and variance σ_h^2, and the ratio of σ_e^2/σ_h^2 was added to the diagonal coefficients of the submatrix for herd-years. Miller et al (1968) reported that the sampling errors of the maximum liklihood estimates were 1.5% smaller than those of least squares estimates. However, the estimates of sampling errors for the maximum liklihood estimates were considered to be biased in this study and also that of Cunningham (1965) because they used the residual variance (σ_e^2) estimated on a least squares model. Thompson (1969) has shown that the reduction in the sum of squares due to a random effect depends on the value of the ratio of residual variance to the variance of the random effect added to the diagonal coefficients of the submatrix for the random effect. He showed that an unbiased estimate of the residual variance in a maximum liklihood analysis can be obtained by the usual procedure but the degrees of freedom for the random effect are not to be subtracted from the total degrees of freedom. A method described by Cunningham and Henderson (1968) for estimating residual variance in the weighted least squares analysis has also been shown to be incorrect by Thompson (1969), as the total degrees of freedom were accounted for incorrectly.

As regards the comparisons of different models for evaluation of sires, the criteria of the product moment correlations between the estimates of sire effects from different models or the rank

correlation are most commonly used. However, the covariances between the estimates of sire effects obtained from two or more independent data sets seems to be a reasonable criterion because an efficient model should give similar estimates of sire effects from the independent data sets. Dempfle and Hagger (1983) compared the efficiency of sire evaluation models using the product moment correlations between sire effects estimated from two data sets using the same model. Henderson (1975a) has described a method for checking the predicted variances on incorrect models assuming that the true model is known. Kennedy and Moxley (1975) used this criteria to compare the two models of sire evaluation: (i) fitting group of sire as an effect ignoring genetic relationships among sires, and (ii) accounting for genetic relationships among sires ignoring groups.

SUMMARY

Since the introduction of BLUP procedure for sire evaluation in early 1970's the herd-year-season effects are traditionally assumed to be fixed, with two or three seasons in most situations. The environmental variation from month to month are usually large, therefore long seasons are not desirable for progeny group comparisons. However, if seasons are taken to be as short as a single month then the herd-year-season subclasses may be very small. The year and seasonal groupings are often completely arbitrary. It could be asked why the year effect is necessary in the model. Could there be a period effect with a length of 6, 18 or 24 months, instead of a year effect? So long as the herd-period-season effects are assumed to be fixed a change in the length of period would not help. But when the season and period effects or just the season

17

effects are assumed to be random, models with periods and seasons of different lengths may be worth investigating.

There are no reports available in which the above mentioned aspects of environmental groupings have been investigated. There are sufficient demonstrations, using examples, of the fact that the sire evaluation models with environmental effects fitted as random give smaller prediction error variances of the estimates of sire effects, but the reports using the field data sets are limited.

TABLE 2.1. Variance (%) contributed by different factors in milk and fat yields.

Source	Herd (H)	Year-Season (YS)	H X YS	Sire (Z)	H X Z	YS X Z	H X YS X Z	Residual
Legates et al (1956): data – all lactations of 17581 daughters by 1260 holstein sires in 5359 herds of the DHIA.								
Milk yield:	30.10			5.90	2.10			61.90
Fat yield :	32.30			6.80	1.20			59.70
Fat % :	15.20			11.80	0.00			73.00
Hickman and Henderson (1955): data – first two lactation records on 3912 Holstein cows in 1094 herds of the USDA-DHIA.								
Milk yield:	29.97	4.04	14.31	6.97	1.71			43.00
Fat yield :	32.87	4.70	14.87	7.26	2.44			37.86
Van Vleck et al (1961): data – 39728 first lactation records of Holsteins from USDA-DHIA.								
Milk yield:	29.14	2.08	6.86	5.89	0.41	-0.75	6.87	49.51
Fat yield :	31.72	2.59	8.50	6.55	0.64	-1.33	3.50	47.84
Allaire and Gaunt (1965): data – 4855 first lactation records in 430 herds by 252 sires (DHIA).								
Milk yield:	33.60			4.70	2.90			58.80
Bereskin and Freeman (1965): data – 39368 records of first three lactations of Holsteins of Holsteins of Iowa-DHIA and DHIR.								
Milk yield:	28.10	2.30	3.20					66.40
Fat yield :	29.20	1.80	4.10					64.90
Harville and Henderson (1967): data – 39644 first lactation records of Holstein from DHIA.								
Milk yield:	26.61	5.13	5.93	8.69	-0.37	-3.11	0.04	57.08
Fat yield :	28.93	4.69	6.72	6.95	-0.03	-2.84	1.04	54.54
Fimland et al (1972b): data – 14612 first lactation records of Israeli-Friesians.								
Milk yield:	25.35			3.91	0.00			70.74

Chapter 3

SEASONAL GROUPING IN A HERD-YEAR-SEASON MODEL OF SIRE EVALUATION

INTRODUCTION

In the sire evaluation procedures used in most breeding schemes
some of the environmental effects are fitted in the model (e.g.
herd, year, season and interactions among these) and some are
precorrected (e.g. month and age of calving). In this method of
environmental corrections it is assumed that the interactions
between the fitted and precorrected effects are unimportant. When
these interactions are an important source of variation or bias in
sire evaluation the joint corrections herd-year-season (Hickman and
Henderson, 1955; Van Vleck, Wadell and Henderson, 1961; Bereskin and
Freeman, 1965; Harville and Henderson, 1967) or age-month (Miller,
Lentz and Henderson, 1970; Mao, Burnside, Wilton and Freeman, 1974;
Cooper and Hargrove, 1982) have been recommended. But the
interactions between the fitted and precorrected effects (e.g.
between herd-year-season and month and age at calving) are of great
concern in sire evaluation. If the effects of herd-year-season and
month of calving were to be fitted as a joint effect in the model it
would result in very small subclasses, making the progeny group
comparisons ineffective.

This study was undertaken to examine the criteria of seasonal
grouping, so that herd-year-season x month of calving interaction

(i.e. herd-year x month within season interaction) is small and the grouping is helpful in giving a smaller variance of sire effects.

MATERIAL AND METHODS

First lactation records on progeny of Holstein-Friesian proven sires in herds participating in the Dairy Progeny Testing Scheme (DPTS) of the Milk Marketing Board of England and Wales (MMB) were extracted from the files supplied by the MMB and edited such that a record must:

(i) have complete information on milk, fat and protein yields,

(ii) pertain to the daughter of a sire which had at least 500 daughters which calved between November, 1972 and October, 1981, and

(iii) have a lactation length between 250 and 305 days (both inclusive), so that it was not necessary to fit the effect of lactation length in the model,

The above editing was done to ease the computations, but there were few records with missing information on either of the milk production traits or with a lactation length of less than 250 days. The qualifying records were on 69 widely used proven sires. From the edited data a total of 49 242 first lactation records were used, and these were split by herd at random into two subsets of manageable size (25 158 and 24 084 records). Both subsets contained records on progeny of all 69 sires and showed similar residual variances.

Least squares analyses were carried out using the LSML76 computer program of Harvey (1977). A "herd-year-month" analysis with fixed effects of herd-year, month of calving, herd-year x month interaction, sire and linear and quadratic regressions on age was undertaken to estimate constants for month of calving (Figure 3.1). Based on these constants the following five seasonal groupings with different trends in production were chosen, in which the grouping (three seasons: December-March, April-July, August-November) currently used in the sire evaluation model of the MMB (G.J.T. Swanson, personal communication) was included.

(i) December-March, April-July and August-November (S-3),

(ii) January-April, May-July and August-December (S-3A). This grouping was chosen because the yields of December calvers were more similar to those of November than January.

(iii) February-July and August-January (S-2A). Both seasons included increasing as well as declining trends in yields. The August-January was considered to be a favourable season in which yields increased gradually until October and then went down, and February-July unfavourable as the yields started going down until May and then increased gradually.

(iv) December-May and June-November (S-2B).

(v) November-May and June-October (S-2C). The seasons of the latter two groupings of unequal size described either a solely increasing (June-October) or declining (November-May) trend.

The different seasonal groupings were compared by the criteria of: (i) smallest herd-year x month within season interaction, (ii) smallest residual variance, (iii) largest effective number of daughters (i.e. diagonal elements of the sire's coefficient matrix after absorption of all enviromental effects in the model with sire as a fixed effect) (iv) smallest variation among months within season. The above criteria were considered to be useful in a herd-year-season model of sire evaluation using records precorrected for the effect of month of calving.

The following model was fitted using different seasonal groupings:

$$Y_{ijklmn} = \mu + HY_{ij} + S_k + M_{kl} + (HYxS)_{ijk} + Z_m + b_1A + b_2A^2 + e_{ijklmn}$$

where:

μ = overall mean

HY_{ij} = fixed effect of the ith herd and the jth year of calving

S_k = fixed effect of the kth season of calving

M_{kl} = fixed effect of the lth month of calving within the kth season

$(HYxS)_{ijk}$ = fixed effect common to progeny in the ith herd, calved

in the jth year and kth season

Z_m = fixed effect of the mth sire (fixed effect since proven sires)

b_1 and b_2 = linear and quadratic regression coefficients on age at

calving (A)

e_{ijklmn} = random error.

The subtraction of the residual sum of squares of the "herd-year-month" analysis from that of the above model gave the sum of squares for herd-year x month within season interaction effect. The degrees of freedom were obtained by subtraction.

Since these data were only of widely used proven sires they were not appropriate for drawing inferences about the effects of different seasonal groupings on alteration in effective number of daughters. Therefore, the effective numbers of daughters were estimated using all the first lactation records of all sires (proven and young) in DPTS herds in a complete year and fitting a model with the fixed effects of herd-year-season and sire. The calculations of effective number of daughters for the S-3 (December-March, April-July and August-November), S-3A (January-April, May-July and August-December) and S-2A (February-July and August-November) groupings were done using the data between August, 1980 and July, 1981 (63 739 records of 3450 sires in 3881 herds), and for S-2B (December-May and June-November) and S-2C (November-May and June-October) between June, 1980 and May, 1981 (63 163 records of 3408 sires in 3873 herds), so that each season in all groupings comprised a complete set of months. Because different sets of data were used, comparisons of groupings were based on the regressions fitted through the origin of the effective number on the actual number of daughters for all sires.

The records used for sire evaluation by the MMB are precorrected for the effects of month and age at calving independently using standard correction factors in all herds of all regions (in Figure 3.2). In this procedure of precorrection of records with the same correction factors it is assumed that there is no region x month of calving interaction. Therefore, region x month of calving and year x month of calving interactions were also examined in two different models to study the seasonal variations from region to region and year to year. In order to partition the sum of squares for region x month interaction a model with fixed effects of region,

24

herd-year-sire within region, month of calving, region x month and herd-year-sire x month interactions, linear and quadratic regressions on age at calving (Table 3.2) was fitted. Similarly the year x month of calving interaction was partitioned using a model with fixed effects of year, herd-sire within year, month of calving, year x month and herd-sire x month interactions, linear and quadratic regressions on age at calving.

The main analysis was conducted (on an untransformed as well as a log scale) using records on progeny of widely used proven sires to enable the estimation of environmental effects more precisely.

RESULTS

Since all inferences drawn from the results of analyses of both subsets of data were the same, full results from only one subset comprising 25 158 records, are presented. However, the main results of the second subset (24 084 records) are also included in Table 3.4. The least squares constants for milk, fat and protein yields and fat and protein contents for the effect of month of calving obtained from the "herd-year-month" analysis are plotted in Figure 3.1. For the three seasons presently used by the MMB the yields of milk, fat and protein of the heifers calving in August-November were observed to be highest followed by those calving in December-March and April-July. The fat contents were highest in August-November and lowest in December-March, and protein contents highest in April-July and lowest in August-November. These results are in agreement with those of Wood (1970) on 6846 heifers from the same breeding scheme for which the highest milk yields were in October-December calvings and the lowest in May-June.

The significance of different factors contributing to variation in yields and composition traits are given in Tables 3.1 and 3.2 ,and these were in general agreement with other reports (Sanders, 1927; Gacula, Gaunt and Damon, 1968; Wood, 1970; Auran, 1973; Lee, 1974; Norman, Kuck, Cassell and Dickinson, 1978; Cooper and Hargrove, 1982). The effects of herd-year, herd-year x season and herd-year x month within season interactions were highly significant (P<0.01) for all traits. The year x month of calving interaction was not significant for milk yield but significant for yields of fat and protein and composition traits.

The weighted means of least squares constants (from both subsets) for milk, fat and protein yields for region-month effects estimated using a model with the joint effect of region and month of calving, and the rest of the effects the same as in the model used for partitioning region x month of calving interaction (in Table 3.2), are shown in Figure 3.2. The seasonal trends tended to be similar for all regions and only small region x month of calving interactions were observed. The effect of region was highly significant (P<0.01) for all traits. The region x month of calving interactions were significant (P<0.05) for analyses of all untransformed yields, but on a log scale were significant only for fat yield. Regional differences were large. However, since the region x month of calving interactions were significant, although small, it would be best to use separate month correction factors for each region for sire analysis to improve comparisons among young bulls which are tested within region at present.

Herd-year x month of calving within season interaction: The mean squares due to season of calving, herd-year x season and herd-year x

26

month within season interactions, and months within each season for milk, fat and protein yields and fat and protein contents, using the different groupings, are given in Table 3.3. Among all analyses with three (S-3 and S-3A) or two seasons (S-2A, S-2B and S-2C) per year with different combinations of equal or unequal number of consecutive months, the S-3 grouping (December-March, April-July and August-November) currently used by the MMB, gave the smallest herd-year x month within season interactions for milk, fat and protein yields and fat and protein contents, however the differences among all groupings were small. As expected, smaller herd-year x month within season interactions were found to be associated with larger herd-year x season interaction effects. The herd-year x month within season interaction effect did not show any association with the between and within season variations. The S-2A season groups showed the largest between season mean squares, more than twice as those in S-3, but there was not much difference in the mean squares for herd-year x month within season interaction effects.

Influence of seasonal groupings on residual variance: The residual variances obtained after fitting the given model using different seasonal groupings, ignoring the herd-year x month within season interaction, from the analyses on an untransformed and a log scale are given in Table 3.4. The residual variance was smallest after fitting a herd-year-month model and largest after fitting only herd-years. Fitting the former is not usually feasible because it results in small environmental groups, whereas the latter does not take account of the distribution of records across seasons, and the herd-year x month interactions are large. Amongst all analyses using a herd-year-season model, the S-3 season groups gave the smallest residual variance for milk, fat and protein yields and fat

and protein contents. The residual variances fitting all seasonal groupings were, however, similar on the untransformed as well as the log scale.

Influence of seasonal groupings on effective number of daughters: The effective number of daughters is important because they enter into the formula for the weights to be given to the estimates of sire effects (Robertson and Rendel, 1954) and into the formula for the prediction error variance in the Best Linear Unbiased Prediction procedure (Henderson, 1973). The regressions of effective numbers (estimated fitting a model with fixed effects of herd-year-season and sire) on the actual number of daughters for all sires (proven and young) fitted through the origin, were estimated to be 0.608, 0.617, 0.638, 0.630 and 0.631 for S-3, S-3A, S-2A, S-2B and S-2C groupings, respectively. These regression coefficients indicated that groupings with two seasons would give 'a slightly larger effective number of daughters than those with three seasons, but the differences were small.

DISCUSSION

In order to account for the seasonal variation in milk production traits it has been recommended in several studies that consecutive calendar months of calving with similar effects on production should be grouped together (Tucker and Legates, 1962; Gaunt, Bartlett and Comstock, 1964; Sargent, Butcher and Legates, 1967). Bereskin and Freeman (1965) suggested the criteria of largest between season, smallest within seasons and smallest residual variances, which is equivalent to grouping months with similar effects. On this basis the S-2A season grouping should have been the best as it showed the

largest between season and smallest within season mean squares, but it gave a larger herd-year x month within season interaction and residual variance than other groupings. In any case the variation between months was small. The S-3A grouping was used to group months with similar effects and gave slightly larger between seasons and smaller within seasons mean squares than S-3, but a larger residual variance. For the purpose of the precorrection of milk records for the effect of month of calving, the smallest herd-year x month within season interaction is considered to be most desirable, but it is not clear how months should be grouped into seasons because the between months within season mean squares did not show any association with the herd-year x month within season interaction effect or residual variance.

If the losses of information due to seasonal groupings were to be taken into account, the effective number of daughters would seem to be a useful criterion for seasonal grouping. The losses of information with any of these five groupings with two or three seasons are negligible, however, because the average number of heifer records available per herd per year is large in DPTS herds, about 15 (MMB, 1984). In a study using the same data, the herd-year-season fixed effects model of sire evaluation with shorter seasons (one or two months) gave larger variances of the estimates of sire effects than it did with longer seasons (Chapter 5). The effective number of daughters would therefore be a more useful criterion where herd sizes are smaller.

29

SUMMARY

Analyses of variance were conducted using two separate random sets of data (25 158 and 24 084 records) on first lactation milk production records on progeny of 69 British Holstein-Friesian proven sires to examine the criteria of seasonal grouping for a sire evaluation model fitting herd-year-season as a fixed effect. Five different seasonal groupings having two or three seasons with equal or unequal number of months and different trends in production within season were chosen.

The smallest herd-year x month of calving within season interaction, smallest residual variance, largest effective number of daughters (diagonal elements of sire's coefficient matrix after absorption of all effects in the model with sire as a fixed effect) and smallest among months within season variation were considered to be the appropriate criteria for determining the effectiveness of the seasonal groupings. None of the groupings examined satisfied all criteria. However, small herd-year x month within season interaction and small within season variation would be desirable if milk records were precorrected for the effect of month of calving in sire evaluation, and the effective number of daughters should be large. On these criteria, the grouping used currently in Britain (three seasons: December-March, April-July and August-November) was considered to be optimal, although a grouping with two seasons: February-July and August-November increased the number of within-group comparisons, the herd-year X month within season interaction was large.

TABLE 3.1. F-ratios for the effects of different factors on milk, fat and protein yields (kg) and fat and protein contents (kg/kg x 100)

Source	d.f.	Milk yield	Fat yield	Protein yield	Fat content	Protein content
Herd-Year-Month model†						
Herd-Year (HY)	3049	6.945**	9.208**	9.043**	4.789**	4.085**
Month of calving (M)	11	34.279**	49.710**	24.831**	13.312**	25.689**
HY X M	7282	1.212**	1.279**	1.290**	1.237**	1.226**
Sire	68	8.503**	7.199**	5.974**	14.679**	16.557**
Regression on age						
at calving: Linear	1	356.599**	425.169**	491.543**	4.233*	22.194**
Quad.	1	64.462**	62.776**	80.503**	0.226	0.397**
Residual m.s.	14745	482 243	690	435	0.088	0.024

Herd-Year-Season model¶ Seasons: (1) December-March, (2) April-July and (3) August-November. Residual m.s. and d.f. same as above

Source	d.f.	Milk yield	Fat yield	Protein yield	Fat content	Protein content
Herd-Year	3232	6.641**	8.801**	8.651**	4.305**	3.920**
Season of calving (S)	2	150.443**	213.988**	112.829**	48.661**	30.986**
HY X S	2617	1.199**	1.302**	1.332**	1.393**	1.399**
HY X M within S	4482	1.121**	1.133**	1.130**	0.999	1.070**
M within S(1)	3	17.004**	20.070**	10.927**	0.480	10.380**
M within S(2)	3	0.054	0.316	1.419	0.964	14.540**
M within S(3)	3	3.940**	3.170*	1.403	0.794	23.874**
Sire	68	10.932**	8.965**	7.538**	19.568**	22.879**
Regression on age						
at calving: Linear	1	560.026**	683.332**	773.286**	10.002**	35.332**
Quad.	1	108.647**	104.123**	125.879**	0.736	0.248

†Effect of month of calving tested against HY X M and all other effects tested against the residual m.s.

¶Effect of season tested against HY X S, HY X S and months (M) within season (S) tested against HY X M within S and all other effects against residual m.s.

TABLE 3.2. F-ratios for the analyses of variance for seasonal variations from region to region and year to year in milk, fat and protein yields (kg) and fat and protein contents (kg/kg x 100)

Source	d.f.	Milk yield	Fat yield	Protein yield	Fat content	Protein content
Region X Month of calving†						
Region (R)	4	20.625**	12.991**	26.056**	21.700**	8.486**
Herd-Year-Sire within region	8000	3.817**	4.762**	4.642**	3.085**	2.957**
Month of calving (M)	11	20.973**	29.023**	15.195**	6.562**	17.893**
R X M	44	1.399*	1.504*	1.402*	1.282	1.775**
Herd-Year-Sire X M within region	7215	1.150**	1.189**	1.203**	1.133**	1.214**
Regression on age at calving: Linear	1	138.412**	165.114**	182.991**	2.343	5.459*
Quad.	1	39.592**	40.917**	47.979**	0.001	0.101
Residual m.s.	9881	463 090	665	419	0.086	0.023

Year X Month of calving¶ (eight years data only)

Year of calving (Y)	7	39.056**	50.246**	58.997**	11.973**	43.622**
Herd-Sire within Y	7195	3.692**	4.516**	4.457**	3.093**	2.801**
Month of calving (M)	11	8.641**	12.749**	7.487**	4.306**	6.000**
Y X M	77	1.223	1.586**	1.303*	1.458**	2.300**
Herd-Sire X M within Y	6338	1.154**	1.196**	1.197**	1.124**	1.186**
Regression on age at calving: Linear	1	122.186**	143.324**	156.059**	1.373	2.102
Quad.	1	41.569**	43.869**	46.166**	0.002	0.007
Residual m.s.	8723	461 341	657	415	0.086	0.023

†Effect of region tested against herd-year-sire, month and region X month against herd-year-sire X month and all other effects against residual m.s.

¶Effect of year tested against herd-sire, month and year X month against herd-sire X month and all other effects against residual m.s.

TABLE 3.3. Mean squares for milk, fat and protein yields (kg^2) and fat and protein contents (kg/kg x 100)^2; due to different factors using different seasonal groupings

Model / Source	d.f.	Milk yield (X10^3)	Fat yield (kg^2)	Protein yield (kg^2)	Fat content (X10^{-2})	Protein content
Herd-Year-Month:						
Herd-Year (HY)	3049	3 349	6 354	3 935	42.15	9.74
Month of calving (M)	11	20 040	43 862	13 937	144.90	75.11
HY X M	7282	585	882	561	10.89	2.92
Sire	68	4 101	4 967	2 599	129.20	39.48
Regression on age						
at calving: Linear	1	171 968	293 391	213 858	37.26	52.92
Quad.	1	31 086	43 319	35,025	1.99	0.95
Residual†	14745	482	690	435	8.80	2.38
Herd-Year-Season (S-3): (1) Dec.-March, (2) April-July and (3) Aug.-Nov.						
Season of calving (S)	2	97 485	217 883	73 922	638.86	110.65
HY X S	2617	648	1 019	655	13.13	3.57
HY X M within S	4482	541	782	492	9.43	2.55
M within S(1)	3	9 192	15 697	5 374	4.53	26.49
M within S(2)	3	29	247	698	9.09	37.11
M within S(3)	3	2 130	2 479	690	7.49	60.93
Herd-Year-Season (S-3A): (1) Jan.-April, (2) May-July and (3) Aug.-Dec.						
Season of calving	2	101 082	222 379	71 858	521.11	189.94
HY X S	2296	631	989	639	12.93	3.46
HY X M within S	4738	550	800	501	9.49	2.59
M within S(1)	3	7 284	11 931	4 844	9.38	18.72
M within S(2)	2	296	120	1 068	6.29	13.10
M within S(3)	4	2 185	3 842	929	28.04	48.03
Herd-Year-Season (S-2A): (1) Feb.-July and (2) Aug.-Jan.						
Season of calving	1	201 168	369 639	138 674	160.07	376.88
HY X S	1473	663	1 047	664	13.03	3.55
HY X M within S	5508	546	805	506	9.83	2.67
M within S(1)	5	3 981	3 192	4 601	78.46	26.95
M within S(2)	5	3 227	9 693	1 483	83.75	39.29
Herd-Year-Season (S-2B): (1) Dec.-May and (2) June-Nov.						
Season of calving	1	52 362	200 289	31 705	1 266.71	145.20
HY X S	1728	691	1 107	706	14.83	3.88
HY X M within S	5371	544	793	502	9.49	2.62
M within S(1)	5	14 544	22 354	9 867	4.63	39.08
M within S(2)	5	11 339	18 160	8 251	6.59	43.31
Herd-Year-Season (S-2C): (1) Nov.-May and (2) June-Oct.						
Season of calving	1	18 217	106 121	13 358	1 163.70	24.89
HY X S	1867	712	1 140	731	15.16	3.77
HY X M within S	5415	541	794	503	9.41	2.63
M within S(1)	6	19 035	33 032	11 251	19.16	70.95
M within S(2)	4	10 016	16 888	7 267	5.27	30.67

† All analyses have the same residual m.s. and d.f.

TABLE 3.4. Residual variances after fitting a herd-year-season model with different seasonal groupings

Seasonal grouping	Residual variance ignoring Herd-year X month within season interaction							
	untransformed					log scale		
	Milk yield	Fat yield	Prot. yield	Fat content	Prot. content	Milk yield	Fat yield	Protein yield
	$(kg^2 \times 10^3)$	(kg^2)	(kg^2)	$(\times 10^{-2})$	$(\times 10^{-2})$		$(\times 10^{-4})$	
Data subset 1: (25 158 records)								
H-Y-M	482	690	435	8.80	2.38	39.83	38.96	34.18
H-Y-S (S-3)	496	712	448	8.95	2.42	41.33	40.45	35.52
H-Y-S (S-3A)	499	717	451	8.97	2.43	41.51	40.68	35.69
H-Y-S (S-2A)	500	721	454	9.08	2.46	41.70	41.04	36.08
H-Y-S (S-2B)	499	717	453	8.99	2.45	41.75	40.97	36.06
H-Y-S (S-2C)	498	718	453	8.97	2.45	41.69	41.05	36.09
H-Y	516	754	477	9.49	2.56	43.32	43.18	38.10
Data subset 2: (24 084 records)								
H-Y-M	473	692	427	8.77	2.29	38.55	38.14	33.06
H-Y-S (S-3)	486	706	438	8.85	2.31	39.62	38.98	34.05
H-Y-S (S-3A)	488	710	440	8.91	2.32	39.77	39.21	34.19
H-Y-S (S-2A)	492	715	444	8.97	2.35	40.13	39.61	34.57
H-Y-S (S-2B)	489	712	442	8.90	2.35	40.03	39.49	34.49
H-Y-S (S-2C)	487	713	441	8.94	2.36	39.93	39.54	34.42
H-Y	510	751	467	9.41	2.47	41.86	41.68	36.52

H = Herd, Y = Year, M = Month and S = Season; for season grouping codes refer to Table 3.3.

FIG. 3.1 Seasonal trends for milk, fat and protein yields and fat and protein contents in British Holstein-Friesian (least squares constants for month of calving). Overall least squares means: milk yield, 4741 kg; fat yield, 183 kg; protein yield, 155 kg; fat content, 3.86 kg/(kgX100); protein content, 3.27 kg/(kgX100).
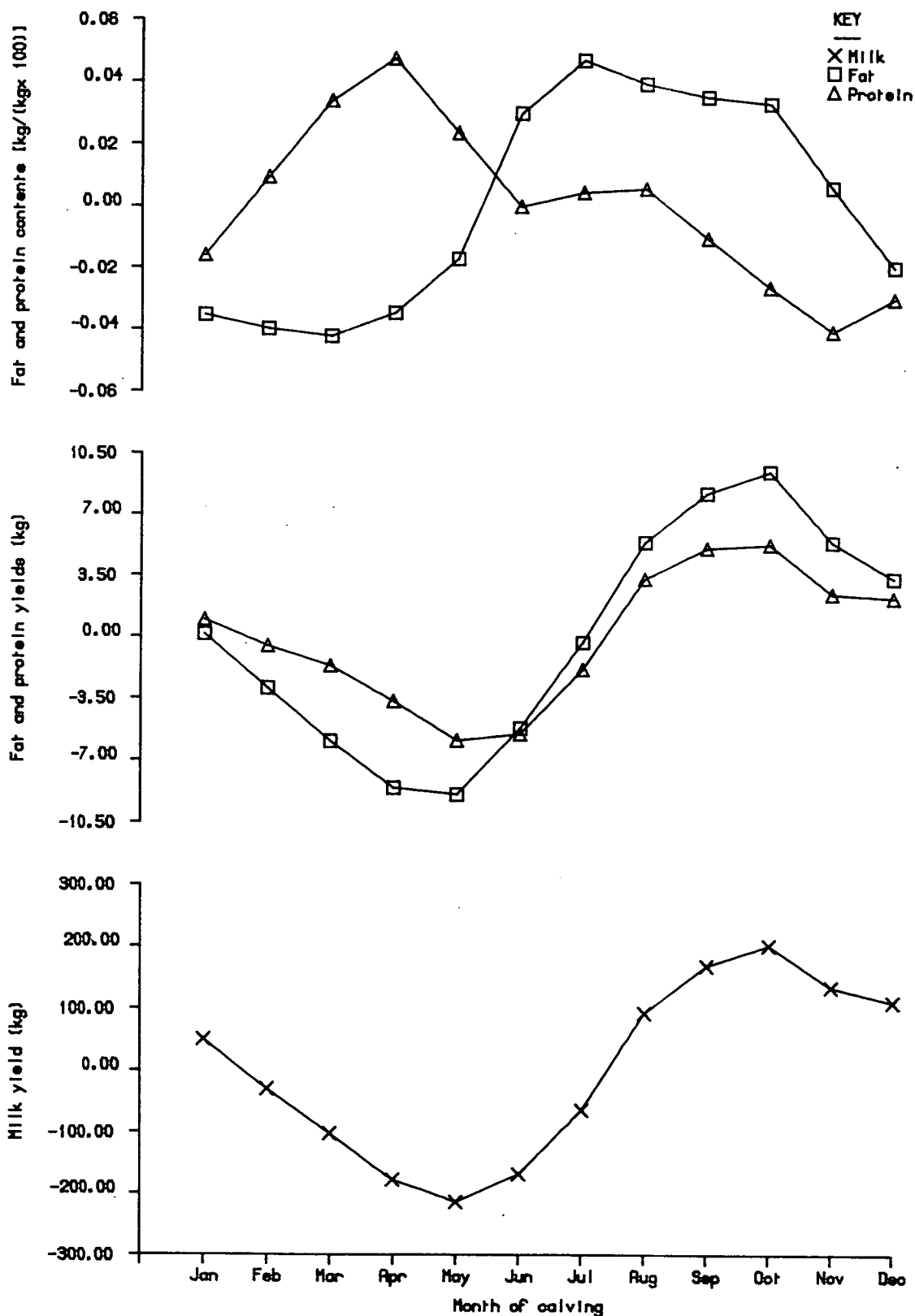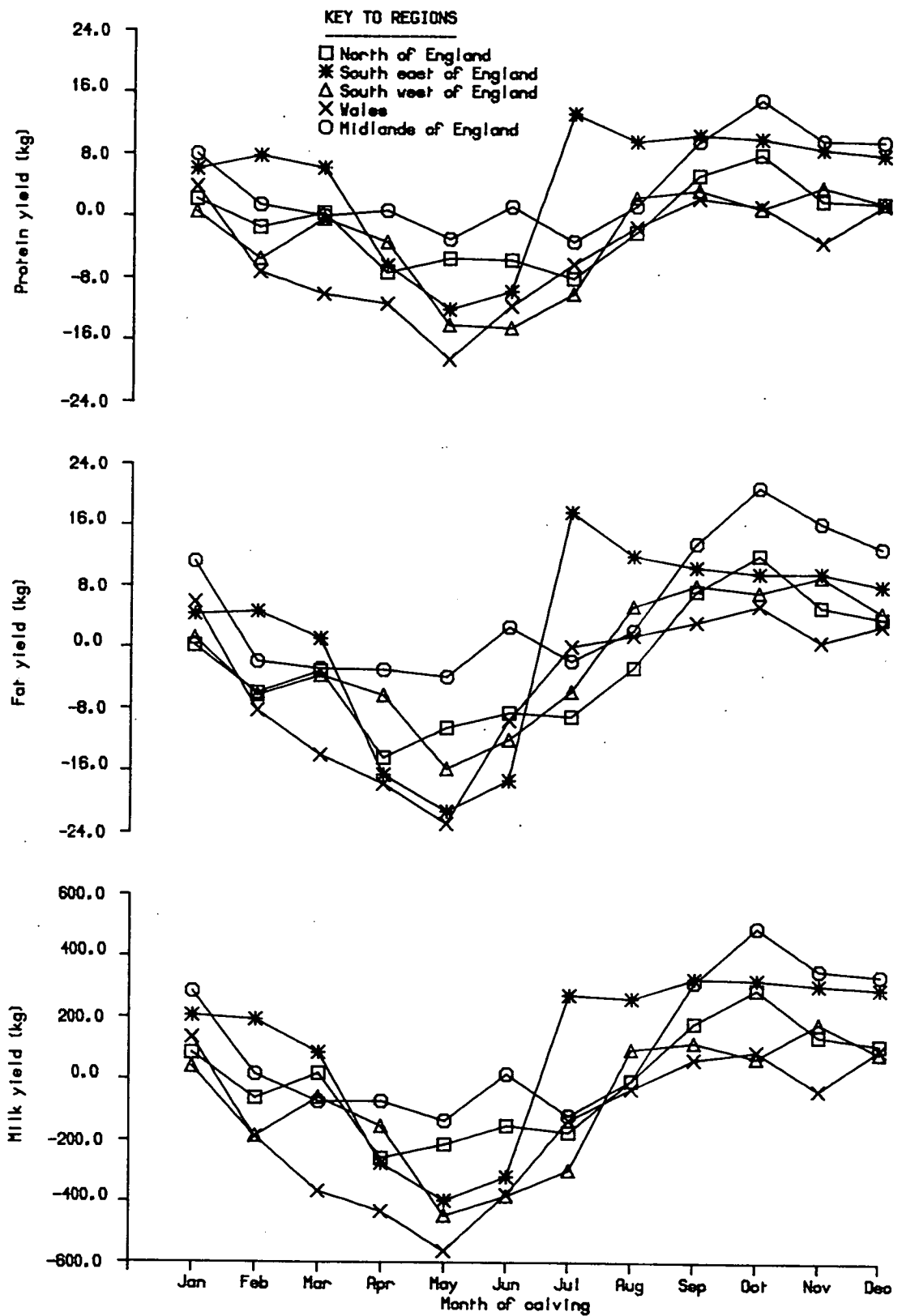
FIG. 3.2 Regional variation in seasonal trends for milk, fat and protein yields
(least squares constants for region-month effects).

KEY TO REGIONS
□ North of England
✳ South east of England
△ South west of England
✕ Wales
○ Midlands of England

Chapter 4

PARTITIONING OF HERD, YEAR, AND SEASON VARIATION IN MILK PRODUCTION

INTRODUCTION

About 40% of the total variance in milk and fat yields is
accounted for by the effects of herd, year and season of calving and
two and three factor interactions among them (Hickman and Henderson,
1955; Van Vleck et al., 1961; Bereskin and Freeman, 1965; Harville
and Henderson, 1967). The proportions of variances accounted for by
different effects fitted in the models of analyses in these studies
are given in Table 4.1. These authors fitted the effects of year
and season of calving as a joint effect, and therefore the two and
three factor interaction variances among herd, year, and season were
not partitioned. Environmental correlations among records in the
same herd, in the same year-season and in the same herd-year-season
have also been reported (Miller, 1964; Thomson and Freeman, 1970;
Arora and Freeman, 1971). There are no reports available, however,
in which variance components for season of variable lengths within
herd-year, have been investigated. Seasons are usually assumed to
be 4 to 7 months long, e.g. Milk Marketing Board (MMB) of England
and Wales (G.J.T. Swanson, personal communication, 1984) and
Northeast Artificial Insemination Sire Comparison (NEAISC) of New
York (Bolgiano et al., 1979). Since the seasonal and management
variations from month to month are generally large, long seasonal
groupings do not seem to be appropriate for progeny group
comparisons. Therefore a model with shorter seasons may be suitable

for estimating the sire effects more precisely.

The main objective of this study was to partition the total variance due to herd-year-season effects such that the variance components could be used in a herd-year-season model of sire evaluation if some of the environmental effects were to be fitted as random effects. Each recording year (December-November) was split into two periods, December-May and June-November. Therefore, the terms "period" and "herd-period-season" have been used instead of year and herd-year-season, respectively. The effects of period and season of calving were fitted with different lengths in order to examine how the variance components would change when the lengths of the period and season changes. Environmental correlations among records within the same herd, the same herd-period and the same herd-period-season were also estimated.

MATERIAL AND METHODS

From the same data as described in Chapter 3 records of the first lactation milk and fat yield on 25158 progeny of 69 widely used proven Holstein-Friesian sires in 832 herds participating in the Dairy Progeny Testing Scheme of the MMB were used. Records of widely used proven sires were used to enable the environmental effects to be estimated more precisely. Since the data were on proven sires, the sire effect was fitted as a fixed effect. The following model with three nested effects was used to estimate variance components for the effects of herd, period within herd and season within period within herd.

$$Y_{ijklm} = \mu + H_i + HP_{ij} + HPS_{ijk} + Z_l + bA + b'A^2 + e_{ijklm}$$

where:

Y_{ijklm} = record of the mth daughter of the lth sire calved at age A in the jth period and the kth season of the ith herd

μ = overall mean

H_i = random effect of the ith herd

HP_{ij} = random effect of the jth period of calving within the ith herd

HPS_{ijk} = random effect of the kth season of calving within the jth period of the ith herd

Z_l = fixed effect of the lth sire (fixed effect since proven sires)

b and b' = linear and quadratic regression coefficients on age at calving A

e_{ijklm} = random error

A least squares analysis of variance was carried out using the LSML76 program of Harvey (1977). Because of the limitations in this program, the analysis of variance (ANOVA) corresponding to the above model had to be synthesized by fitting several models separately, and combining the mean squares (m.s.), degrees of freedom (d.f.) and coefficients (k-values) of variance components. The procedure for constructing an ANOVA for three nested effects model using m.s., d.f. and k-values from two nested effects analyses is given in Table 4.2. All k-values except one (k_2) were available from two nested effects analyses. The k_2 was calculated as described by Mahamunulu (1963). This procedure (Table 4.2) was verified by analysing example data with two herds, using MTY option 5 in the LSML76 program.

The data were precorrected for the effects of year and month of calving using least squares constants estimated from the same data. The constants for year effects were estimated fitting a model with the effects of year, herd-month within year and sire with age at

calving as a covariable (linear and quadratic regressions). Constants for month of calving were obtained using a similar model but with the effects of month and herd-year within month. This method was a combination of Henderson's methods 2 and 3 (Henderson, 1953), because some of the fixed effects were precorrected for, and some fixed effects fitted in the model.

Four environmental groupings, given in Table 4.3, were used. The variance components for season within period within herd are directly comparable between groupings only when the period effects are the same. Similarly, components for period within herd are comparable when season effects are the same. Grouping A (period of 12 months and season 4 months) is presently used by the MMB in their sire evaluation model, with herd-period-season fitted as a fixed effect. Grouping C (period of 12 month and season one month) was chosen to fit shorter seasons and to compare it to grouping A. Groupings B (period of 6 months and season 2 months) and D (period of 6 months and season one month) also gave comparisons of variance components for seasons within period within herd, with different lengths. The component of variance for the effects of period within herd in grouping D is comparable to the period component in grouping C.

The formulae for estimating environmental correlations have been given by Miller (1964) and Thomson and Freeman (1970), using a model with the effects of herd, year-season, sire, and daughter all fitted as random effects. In their models they ignored the herd x year-season interaction. This herd x year-season interaction effect should have been included in their model, however, for estimating the correlation among records within the same herd-year-season.

40

The following formulae were used in this study for calculating the enviromental correlations from the variance components estimated using the given model.

same herd : $\sigma_h^2/(\sigma_h^2+\sigma_{hp}^2+\sigma_{hps}^2+\sigma_e^2)$

same herd-period : $(\sigma_h^2+\sigma_{hp}^2)/(\sigma_h^2+\sigma_{hp}^2+\sigma_{hps}^2+\sigma_e^2)$

same herd-period-season: $(\sigma_h^2+\sigma_{hp}^2+\sigma_{hps}^2)/(\sigma_h^2+\sigma_{hp}^2+\sigma_{hps}^2+\sigma_e^2)$

The quantities σ_h^2, σ_{hp}^2, σ_{hps}^2 and σ_e^2 are the variances due to herd, period within herd, season within period within herd and residual, respectively. For calculating the correlations assuming the herd as a fixed effect, the σ_h^2 component would not be included in the above formulae. This would cause the "same herd" correlation, i.e. the correlation amongst records in the same herd, to be zero.

RESULTS

The variance components estimated fitting period and season with different lengths are given in Table 4.3.

Using grouping A (period of 12 months and season 4 month) the proportions of variances due to herd, period within herd, season within period within herd, and residual were; milk yield: 0.31, 0.06, 0.05 and 0.58 respectively, and fat yield: 0.35, 0.08, 0.07 and 0.50. These proportions with different groupings were similar. The residual variance was, however, smallest when each season comprised one month (groupings C and D). It was also noticed that the residual variance was proportionately less for fat yield than

for milk yield. The results of other studies given in Table 4.1 also showed similar patterns in which residual variances for fat yield were about 2 to 5% smaller than those for milk yield.

The environmental correlations assuming herd as either a fixed or random effect and period within herd and season within period within herd as random effects, are given in Table 4.4. Using both sets of correlations, the comparisons between groupings A and C, and B and D, showed that correlations among records within the same herd-period-season were higher when seasons were shorter, but the differences were small. Comparisons between groupings C and D also showed that correlation within herd-period were higher when period comprised 6 months, as opposed to 12 months.

DISCUSSION

Although the models used were slightly different, the proportions of the components of variance for herd and residual were in agreement with those of Hickman and Henderson (1955), Van Vleck et al (1961), and Harville and Henderson (1967). The residual variances for milk and fat yields reported by Bereskin and Freeman (1965) were about 10% higher than the estimates found in this study and also of those of Hickman and Henderson (1955), Van Vleck et al (1961), and Harville and Henderson (1967). These differences were presumably due to the fact that Bereskin and Freeman did not fit the effect of sire. In the model used in this study, the effects of period within herd and season within period within herd, comprise the variances due to the two and three factor interactions among herd, period and season. These effects were considered to be

important sources of variation as they accounted for about one-third of the total variance due to the herd-period-season effects (excluding overall effects of period and season of calving).

The environmental correlations found in this study were considerably higher than those of Thomson and Freeman (1970) and Arora and Freeman (1971). Thomson and Freeman reported correlations (for milk yield) of 0.226 within the same herd and 0.328 within the same herd-year-season. Correlations within the same herd, and in the same herd-year-season, reported by Arora and Freeman, were 0.198 and 0.357 respectively for milk yield, and 0.216 and 0.381, respectively, for fat yield. The estimates of these two studies seem to be biased downward because they ignored herd x year-season interaction effect in their model. Correlations among records in the same herd and in the same herd-year-season, computed from the proportions of variance components reported by Hickman and Henderson (1955), Van Vleck et al (1961), and Harville and Henderson (1967) (Table 4.1), were in the ranges of 0.27 - 0.30 and 0.38 - 0.48 respectively for milk yield and 0.29 - 0.33 and 0.40 - 0.52, respectively, for fat yield.

The differences in the environmental correlations among records within the same herd-period-season with a season of either 1 or 2 or 4 months (Table 4.4) were small. It was concluded that seasons shorter than 4 months in the sire evaluation model of the MMB (herd-year-season fixed) would not be very helpful, because shorter seasons will split records into small subclasses with a few or a single record, and these small subclasses will contribute a little or nothing to progeny group comparisons. However, if the effects of period within herd and season within period within herd are assumed

to be random then there are two possible models for sire evaluation: (i) herd fixed and herd-period-season random, (ii) herd fixed, herd-period and herd-period-season random. The latter model takes account of the fact that the correlations among records within the same herd-period-season are higher than those within the same herd-period. Therefore, a sire evaluation model with herd as a fixed effect and herd-period and herd-period-season as random effects may be more appropriate than a model with herd fixed and herd-period-season random (see Chapter 5).

SUMMARY

Variances in milk and fat yields due to herd-period-season effects (period either 6 or 12 months, and season either 1 or 2 or 4 months) were partitioned fitting a three nested effects model with herd, period within herd and season within period within herd as random nested effects, sire a fixed effect and linear and quadratic regressions on age at calving. The overall effects of period and season of calving were regarded as fixed effects and were removed by precorrecting records using least squares constants estimated from the same data. Environmental correlations within herd, herd-period and herd-period-season were also estimated for all period and season groupings. Records of first lactation milk and fat yields on 25158 progeny of 69 widely used proven Holstein-Friesian sires in 832 herds in England and Wales were used.

The variance components for the effects of herd, period within herd, season within period within herd and residual accounted for 31, 6, 5 and 58% of the total variance in milk yield, and 35, 8, 7 and 50% of the variance in fat yield, respectively, using a period

of 12 months and a season of 4 months. Differences amongst correlations within the same herd-period-season, fitting seasons of different lengths, were small. It was therefore concluded that shorter seasons in a herd-period-season fixed effect model of sire evaluation would be of no advantage.

TABLE 4.1. Variance (%) contributed by different factors in milk and fat yields.

Source	Herd (H)	Year-Season (YS)	H X YS	Sire (Z)	H X Z	YS X Z	H X YS X Z	Residual
Hickman and Henderson (1955)								
Milk yield:	29.97	4.04	14.31	6.97	1.71			43.00
Fat yield :	32.87	4.70	14.87	7.26	2.44			37.86
Van Vleck et al (1961)								
Milk yield:	29.14	2.08	6.86	5.89	0.41	-0.75	6.87	49.51
Fat yield :	31.72	2.59	8.50	6.55	0.64	-1.33	3.50	47.84
Bereskin and Freeman (1965)								
Milk yield:	28.10	2.30	3.20					66.40
Fat yield :	29.20	1.80	4.10					64.90
Harville and Henderson (1967)								
Milk yield:	26.61	5.13	5.93	8.69	-0.37	-3.11	0.04	57.08
Fat yield :	28.93	4.69	6.72	6.95	-0.03	-2.84	1.04	54.54

TABLE 4.2. Constructing ANOVA for estimating variance components in a three nested effects structure using information from two nested effects analyses with LSML76 computer program of Harvey (1977).

Source	D.F.	M.S.	E.M.S.

Analysis 1: H = herd, HPS = period-season within herd

Source	D.F.	M.S.	E.M.S.
H	DF(h)	MS(h)	$\sigma_e^2 + k_{12}\sigma_{HPS}^2 + k_{13}\sigma_H^2$
HPS	DF(HPS)	MS(HPS)	$\sigma_e^2 + k_{11}\sigma_{HPS}^2$
Residual	DF(e)	MS(e)	σ_e^2

Analysis 2: HP = herd-period, HPS = season within herd-period

Source	D.F.	M.S.	E.M.S.
HP	DF(HP)	MS(HP)	$\sigma_e^2 + k_{22}\sigma_{HPS}^2 + k_{23}\sigma_{HP}^2$
HPS	DF(hps)	MS(hps)	$\sigma_e^2 + k_{21}\sigma_{HPS}^2$
Residual	DF(e)	MS(e)	σ_e^2

Analysis 3: H = herd, HP = period within herd,

Source	D.F.	M.S.	E.M.S.
H	DF(h)	MS(h)	$\sigma_E^2 + k_{32}\sigma_{HP}^2 + k_{33}\sigma_H^2$
HP	DF(hp)	MS(hp)	$\sigma_E^2 + k_{31}\sigma_{HP}^2$
Residual	DF(E)	MS(E)	σ_E^2

Final ANOVA[†]: H = herd, HP = period within herd, HPS = season within period within herd,

Source	D.F.	M.S.	E.M.S.
H	DF(h)	MS(h)	$\sigma_e^2 + k_{12}\sigma_{hps}^2 + k_{32}\sigma_{hp}^2 + k_{33}\sigma_h^2$
HP	DF(hp)	MS(hp)	$\sigma_e^2 + k_2\sigma_{hps}^2 + k_{31}\sigma_{hp}^2$
HPS	DF(hps)	MS(hps)	$\sigma_e^2 + k_{21}\sigma_{hps}^2$
Residual	DF(e)	MS(e)	σ_e^2

$$k_2 = [k_{11}\{DF(HPS)\} - k_{21}\{DF(hps)\}]/[\{DF(HPS)\} - \{DF(hps)\}]$$

[†] 1. The M.S. and D.F. indicated with small letters in parentheses and some k-values from the first three analyses were used to construct ANOVA. k_{13} and k_{33} are the same. The first subscript of k-values indicate the analysis they come from,
2. Only the residual variance indicated with a small subscript is relevant to final ANOVA, and those with subscripts using bold letters are appropriate for two nested effects models in respective analysis,
3. Effects of sire (fixed) and age as a covariable fitted in all analyses, and effects of period and season of calving removed by precorrecting records.

47

TABLE 4.3. Variance components for milk and fat yields with different environmental groupings and subclasses.

		Grouping			
		A	B	C	D
Herd	(h):	herd	herd	herd	herd
Period	(p):	12 month	6 month	12 month	6 month
Season	(s):	4 month	2 month	1 month	1 month

Number of Subclasses:	A	B	C	D
Herd	832	832	832	832
Herd-period	3233	4962	3233	4962
Herd-period-season	5852	7974	10344	10344

Milk yield ($kg^2 \times 10^3$)

	A	B	C	D
σ_h^2	264 (0.31)	268	264	268
σ_{hp}^2	52 (0.06)	73	66	78
σ_{hps}^2	46 (0.05)	30	46	30
σ_e^2	497 (0.58)	488	482	482

Fat yield (kg^2)

	A	B	C	D
σ_h^2	500 (0.35)	509	500	509
σ_{hp}^2	113 (0.08)	161	144	171
σ_{hps}^2	93 (0.07)	50	85	49
σ_e^2	714 (0.50)	699	690	690

NB: As the effects of period and season were removed by precorrecting records σ_{hp}^2 comprises variance due to HxP interaction, and σ_{hps}^2 variance due to HxS, PxS and HxPxS interactions. σ_h^2 and σ_e^2 are herd and residual variances, respectively.

Figures in parentheses are proportions of total variance.

TABLE 4.4. Environmental correlations.

	Environmental grouping			
	A	B	C	D
Period: Season:	12 months 4 months	6 months 2 months	12 months 1 month	6 months 1 month

Assuming herd as a fixed effect, and period within herd and season within period within herd as random effects

Milk yield:

Same herd-period	0.09	0.12	0.11	0.13
same herd-period-season	0.16	0.17	0.19	0.18

Fat yield:

Same herd-period	0.12	0.18	0.16	0.19
Same herd-period-season	0.22	0.23	0.25	0.24

Assuming herd, period within herd and season within period within herd as random effects

Milk yield:

Same herd	0.31	0.31	0.31	0.31
Same herd-period	0.37	0.40	0.38	0.40
same herd-period-season	0.42	0.43	0.44	0.44

Fat yield:

Same herd	0.35	0.36	0.35	0.36
Same herd-period	0.43	0.47	0.45	0.48
Same herd-period-season	0.50	0.51	0.51	0.51

NB: Overall effects of period and season of calving removed by precorrecting records.

Chapter 5

DAIRY SIRE EVALUATION FITTING SOME OF THE HERD-YEAR-SEASON EFFECTS AS RANDOM

INTRODUCTION

In the sire evaluation model used in most breeding schemes herd-year-season is fitted as a fixed effect and it is the main environmental grouping, with two or three seasons, within which progeny groups are compared. For example, the Northeast Artificial Insemination Sire Comparison (NEAISC) of New York uses two seasons: December-April and May-November (Bolgiano et al., 1979), and the Milk Marketing Board (MMB) of England and Wales uses three seasons: December-March, April-July and August-November (G.J.T. Swanson, personal communication, 1984). The effect of herd-year-season is fitted as a fixed effect in order to allow corrections for the possible bias due to non-random use of sires across herds, because some breeders tend to use bulls of a higher genetic merit. However, it is reasonable to assume that breeders use bulls of a similar genetic merit in any one year and also presumably they do not tend to use any one of the bulls in a particular season. If this breeding practice applies, although the herd effects would be fixed, the effects of year within herd and season within year within herd could be assumed to be random, and extra information could be recovered on progeny group comparisons from herd-year and herd-year-season blocks (Cunningham and Henderson, 1966; Patterson

50

and Thompson, 1971). Year is generally split into two or three seasons and the cut off points are decided by the date of calving. A few hours difference in calving may result in distributing animals into different herd-year-season contemporary groups. When seasons are taken to be 4 to 7 months long *e.g.* (MMB and NEAISC), a large number of animals may have contemporaries which perform under a completely different environment, so shorter season grouping should be more efficient. In this study, therefore, smaller seasonal groupings were considered along with the suitability of different models using different groupings. Year was split into two shorter periods in one of the groupings and, therefore, the terms "herd-period" and "herd-period-season" have been used instead of herd-year and herd-year-season, respectively.

Cunningham (1965) and Henderson (1973, 1975a, 1975b), in simple examples, and Miller et al (1968) using field data, have compared sire evaluation models in which a single environmental factor was fitted as a fixed or random effect. They showed that a random effects model gave smaller sampling errors of the estimates of sire effects. However, no reports are available in which environmental factors such as herd, period within herd, and season within period within herd are fitted in a nested structure, with one or more of the nested factors fitted as random effects.

This study was undertaken to compare sire evaluation models with alternative environmental groupings and with some environmental effects fitted as random.

MATERIAL AND METHODS

The data comprised nine years of records on first lactation fat yields of 49242 progeny of 69 widely used proven Holstein-Friesian sires in 1628 herds participating in the Dairy Progeny Testing Scheme (DPTS) of the MMB and as described in Chapter 3. The number of daughters per sire ranged between 92 and 3681 with an average of 714. As only the records on progeny of widely used sires were extracted about 50% of the first lactation records were discarded, resulting in an average of 7.8 heifer records per herd per year whereas in the DPTS herds the average number of heifer records is 15 per herd per year (MMB, 1984). The data on widely used proven sires were used so that it was possible to split them into several random subsets in order to compare different models using empirical covariances of the estimates of sire effects between subsets. These data were split by herd at random into 8 subsets. With this, all records of a herd were included in any one of the subsets. Each subset contained records of all sires. The data were precorrected for the effects of year and month of calving using least squares constants estimated from the same data (Chapter 4). Precorrections were done mainly to avoid having to fit the crossclassified fixed effects of period and season of calving. The variance components used in the weighted least squares estimation of sire effects were also estimated from precorrected records, as described in Chapter 4. The ratios of residual variance to the variance of particular environmental factors and environmental groupings used are given in Table 5.1. In grouping A and C a period comprised 12 months (December-November) whereas in grouping B it was split into two periods each comprising 6-months: December-May and June-November. The seasons were of either 1 or 2 or 4 months.

52

The following three models were compared using each grouping.

Model 1 : $Y_{ijklm} = HPS_{ijk} + Z_1 + bA + b'A^2 + e_{ijklm}$

Model 2 : $Y_{ijklm} = HP_{ij} + S_{ijk} + Z_1 + bA + b'A^2 + e_{ijklm}$

Model 3 : $Y_{ijklm} = H_i + P_{ij} + S_{ijk} + Z_1 + bA + b'A^2 + e_{ijklm}$

where:

Y_{ijklm} = record of the mth progeny of the lth sire calved at age A

in the jth period and the kth season of the ith herd,

HPS_{ijk} = joint fixed effect of the ith herd, the jth period and the kth

season of calving,

HP_{ij} = joint fixed effect of the ith herd and the jth period of calving,

H_i = fixed effect of the ith herd,

S_{ijk} = random effect of the kth season of calving in the jth period and in

the ith herd,

P_{ij} = random effect of the jth period of calving in the ith herd,

Z_1 = fixed effect of the lth sire (fixed effect since proven sires),

b and b' = linear and quadratic regression coefficients on age at calving A.

e_{ijklm} = random error.

The mixed model equations (MME) for model 3 (herd fixed, herd-period and

herd-period-season random) are given below in matrix notation:

$$
\begin{bmatrix}
H'H & H'P & H'S & H'Z & H'C \\
P'H & P'P+ \lambda I & P'S & P'Z & P'C \\
S'H & S'P & S'S+ \alpha I & S'Z & S'C \\
Z'H & Z'P & Z'S & Z'Z & Z'C \\
C'H & C'P & C'S & C'Z & C'C
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3 \\
u \\
b_4
\end{bmatrix}
=
\begin{bmatrix}
H'Y \\
P'Y \\
S'Y \\
Z'Y \\
C'Y
\end{bmatrix}
$$

wher:

Y = vector of fat yield records

H = design matrix for herds

P = design matrix for herd-periods

S = design matrix for herd-period-seasons

Z = design matrix for sires

C = matrix of information on covariables of age and square of
age at calving

b_1 = vector of unknown fixed effects for herds

b_2 = vector of unknown random effects for herd-periods

b_3 = vector of unknown random effects for herd-period-seasons

b_4 = vector of unknown linear and quadratic regression coefficients
of fat yield on age at calving

u = vector of unknown fixed effects for sires

λ = ratio of residual variance to the variance due to
herd-periods, σ_e^2/σ_{hp}^2

α = ratio of residual variance to the variance due to

herd-period-seasons, $\sigma_e^2/\sigma_{hps}^2$

I = identity matrix

The sires' equations were solved in the following order:

$M = [I - S(S'S+\alpha I)^{-1}S']$ ..(adjustment for random effect of herd-period-seaso

$Q = [M - MP(P'MP+\lambda I)^{-1}P'M]$ ..(adjustment for random effect of herd-period)

$R = [Q - QH(H'QH)^{-1}H'Q]$ ... (adjustment for fixed effect of herd)

After processing the data of each herd the equations for age

and sire were accumulated and then sire equations were solved by

absorbing age equations as follows:

$T = [R - RC(C'RC)^{-1}C'R]$

$u = [Z'TZ]^{-1}[Z'TY]$

The residual mean squares (RMS) were estimated by the procedure

described by Thompson (1969), which was a correction to the method

of Cunningham and Henderson (1968) for estimating the residual

variance in the weighted least squares procedure,

$RMS = [Y'TY - Y'TZ (Z'TZ)^{-1}Z'TY]/[N - \text{rank(fixed effects)}]$;

where, N = total number of records,

The variances of sire effects were estimated by the procedure given

by Henderson (1973), as the product of the RMS and the corresponding

diagonal element of the inverse of Z'TZ.

The MME for model 2 (herd-period fixed and herd-period-season random) can be obtained from the MME of model 3 by deleting the row and column pertaining to herd (H) equations and not adding λI to the P'P submatrix of the coefficient matrix. Similarly, the MME for model 1 (herd-period-season fixed) can be obtained by deleting the rows and columns pertaining to herd (H) and herd-period (P) equations and not adding αI to the S'S submatrix.

Criteria of comparison of models

The product moment correlations between sire effects from two random subsets have been used (Dempfle and Hagger, 1983; Hagger and Dempfle, 1983) as a criterion to compare different models of sire evaluation. This procedure ignores the fact that the estimates of sire effects have unequal variances when data are unbalanced. However, in addition to the best linear unbiased predictor (BLUP) properties (Henderson, 1973, 1975b), the following properties of a sire evaluation model are also desirable: It should (i) give the minimum variance of sire effects, (ii) be able to make herd-period-season corrections such that the sire effects are similar from all independent subsets, (iii) give estimated variances that are close to their expected values. Based on these properties the following three criteria were chosen to compare different models. (a) Relative values of estimated variances of sire effects; (b) Empirical variance: empirical variances for each sire were calculated using the formula given below,

Empirical variance $= [\sum_{j=1}^{P} (u_{ij} - \bar{u}_i)^2]/(P-1)$

where:

$$\bar{u}_i = (\sum_{j=1}^{P} u_{ij})/P$$

u_{ij} = sire effect for the ith sire from the jth subset of data,

P   =   number of subsets of data

(c) Ratio of empirical to estimated variance:

Ratio $= \sum_{j=1}^{P} [(u_{ij} - \bar{S}_i)^2/V_{ij}]$

where:

$$\bar{S}_i = [\sum_{j=1}^{P} (u_{ij}/V_{ij})]/[\sum_{j=1}^{P} (1/V_{ij})]$$

V_{ij} = estimated variance of u_{ij} for the specified model,

In the above formulae $\bar{u}_i$ is the simple mean of sire effects for a sire from all subsets and $\bar{S}_i$ is the weighted mean. The expected value of the empirical variance is the variance due to sire, whereas the expectation of the ratio of empirical to estimated variance is (P-1). The empirical variances and the ratios were calculated for each sire and then averaged. As the empirical variance decreases the efficiency of the model increases. Similarly, a value of the ratio of empirical to estimated variance close to unity is desirable. The methods (a) and (b) have algebraic similarity and may give similar inferences about comparison of models. However, the ratio method was considered to be a logical criterion because it takes account of the unequal variance of sire effects from each subset, and could be useful especially when the differences in the number of progeny between subsets are large. The ratio method and the product moment correlations (between sire effects from two subsets) have one to one correspondence when sire effects have equal

56

variances.

In order to contrast different methods for comparing models, the intraclass correlations among sire effects from all subsets were also estimated. The efficiency of different models was illustrated in terms of the the effective number of daughters they would require to give the equivalent variances of sire effects. The effective numbers were defined as the corresponding diagonal elements of the sires' coefficient matrix after absorption of all environmental effects in the herd-period-season fixed effect model. The predicted effective numbers for the equivalent variances of the estimates of sire effects were calculated using a regression of effective numbers on the 'inverse of estimated variances' of sire effects from each model.

RESULTS

The relative values of the estimated variances of sire effects using different models and environmental groupings are given in Table 5.2. Compared to model 1 (herd-period-season fixed), model 2 (herd-period fixed and herd-period-season random) gave 16, 17 and 37% smaller variances, and the model 3 (herd fixed, herd-period and herd-period-season random) 32, 40 and 49% smaller variances using groupings A, B and C, respectively. Comparisons between models 2 and 3 demonstrated what would be gained if the period effects within herd were assumed to be random. These results (Table 5.2) showed that model 3 was more efficient than model 2 by about 29 and 19% using periods of 6 months (grouping B) and 12 months (groupings A and C), respectively.

When different enviromental groupings were compared within model it was observed that the variances of the estimates of sire effects using model 1 (herd-period-season fixed) were larger by 14% with a season of 2 months (grouping B) and 36% with a season of one month (grouping C), than those of a season of 4 months (grouping A). Although shorter seasons are desirable for comparing progeny groups, these results showed that with a herd-period-season fixed effect model the losses of information were substantial. This is because the herd-period-season subclasses which contain a single or few records, or subclasses in which all or almost all records are of the same sire, contribute nothing or little to progeny group comparisons. These losses could be avoided using the random effects model.

The trends in the estimates of the empirical variances (Table 5.3) and intraclass correlations among sire effects (Table 5.4) were quite similar to those of the relative values of estimated variances for all models, leading to the same inferences about comparison of models - model 3 being the best followed by models 2 and 1.

The ratios of empirical to estimated variances, each of them divided by the expected value of 7 (i.e. number of subsets minus one), are given in Table 5.3. All models underestimated the variances of sire effects by a similar proportion of about 18%. It implies that all three models were equally accurate.

The product moment correlations between the sire effects estimated using different models for all enviromnental groupings are given in Table 5.4. All correlations were close to unity. Table 5.5 shows the effective number of daughters required by different

models to give the equivalent variances of the estimates of sire effects.

DISCUSSION

Based on the criteria of the relative values of estimated variance, empirical variance and intraclass correlations among sire effects, model 3 (herd fixed, herd-period and herd-period-season random) was observed to be the best followed by models 2 and 1. It was concluded that the random effects models were more efficient than the herd-period-season fixed effect model, as the averages of the empirical variances were smaller (Table 5.3) and also required a smaller number of progeny to yield the equivalent variance of sire effects (Table 5.5). While examining the sire effects for each sire from all subsets, it was noticed that, although most of the sires had a similar number of progeny in all subsets, a few sires had uniform numbers in most subsets but smaller numbers in one or two of the subsets. The sire effects from such subsets (with small numbers) were much larger than the sire effects from the rest of the subsets. In these cases the empirical variances were exceptionally large. However, the ratios of empirical to estimated variances would not be so sensitive because the estimated variance of sire effects based on small numbers would also be larger.

The product moment correlations between sire effects from different models (Table 5.4) were close to unity, indicating that using random effects models (models 2 and 3) the same sires would be selected as would be by using the herd-period-season fixed effect model. The random effects models were considered to be more

efficient, however, because they require a considerably smaller
effective numbers of daughters (Table 5.5) than the
herd-period-season fixed effect model to give the equivalent
variances of the estimates of sire effects. This implies that the
random effects models could allow testing of a larger number of
young bulls with the same amount of resources. Consequently, a
higher intensity of selection could be applied.

If the breeding practices are such that periodic trends
(environmental or genetic) existed within herds, however, then the
model 2 in which herd-period is fitted as a fixed effect, should be
preferred over model 3. Using the criterion of empirical variances,
a model with herd-year effects as fixed and herd-year-month random
(model 2 with grouping C) was observed to be 37.3% more efficient
than the herd-year-season fixed effects model (i.e. model 1 with
grouping A), currently used by the MMB. The empirical variances for
each sire have been plotted in Figure 5.1 for model 2 with grouping
C against those for model 1 with grouping A, showing that the
empirical variances for most sire were much smaller on the former
model. Alternatively, a model similar to model 3 with herd-period
(period of 12 months) as a fixed effect, and herd-period-season
(season of 3 or 4 months) and herd-period-month as random effects
could be suitable.

Random effects models could be especially useful where herds are
small. For instance, in the sire evaluation model of the dairy
cattle breeding schemes in Switzerland the effect of herd production
level-region-year-season is the main environmental group for
Braunvieh, and herd production level-region-year for Black-and-White
cattle (Hagger et al., 1984). The herd effects are ignored because

60

herd-year-season size is small. Fitting the effect of herd as a random effect within herd production level-region-year-season would presumably be more useful than ignoring it completely. Alternatively, a model in which the effect of herd production level-region (HR) is a fixed effect, herd-year within HR is a random effect and month within herd-year is also a random effect, could be worth investigating. The overall effects of year and month of calving could be fitted in the model as fixed effects or precorrected.

SUMMARY

Three models of sire evaluation using different environmental groupings were compared. Effects fitted were herd, period (either 6 or 12 months) within herd, season (either 1 or 2 or 4 months) within period within herd, sire and linear and quadratic regressions on age at calving. Models differed in fitting (1) the effect of herd-period-season fixed, or (2) herd-period fixed and herd-period-season random, or (3) herd fixed, herd-period and herd-period-season random. The overall effects of period and season of calving were regarded as fixed, and were removed by precorrecting the records using least squares constants estimated from the same data. Records of first lactation fat yield on 49242 progeny of 69 widely used proven Holstein-Friesian sires in 1628 herds in England and Wales were used, and sire effects were fitted as fixed.

Compared to model 1, model 2 required about 4/5ths and model 3 2/3rds of the effective number of daughters (i.e. diagonal elements of the sire's coefficient matrix after absorption of all

environmental effects in the herd-period-season fixed effect model) to give the equivalent variance of the estimates of sire effects. However, all models would seem to give similar ranking of sires. Using the random effects models the relative advantages, in terms of smaller variances of sire effects, increased as the sizes of herd-period-season subclasses decreased.

When herd-period-season is fitted as a fixed effect there may be considerable loss of information because the herd-period-season subclasses which contain a single or few records, or subclasses in which all or almost all records are of the same sire, contribute nothing or little to the progeny group comparisons. The random effects models could avoid these losses, and were considered to be useful especially where herds are small.

TABLE 5.1. Ratios of residual variance ($\sigma_e{}^2$) to different environmental variance components for fat yield with different environmental groupings and subclasses.

	Environmental Grouping		
	A	B	C
Herd (h):	Herd	Herd	Herd
Period (p):	12 months	6 months	12 months
Season (s):	4 months within p	2 months within p	calendar month within p

Ratios:

	A	B	C
$\sigma_e{}^2/\sigma_{hp}{}^2$	6.29	4.33	4.78
$\sigma_e{}^2/\sigma_{hps}{}^2$	7.68	14.07	8.07

Average size of subclasses:

	A	B	C
Herd	30.2	30.2	30.2
Herd-period	7.8	5.1	7.8
Herd-period-season	4.3	3.2	2.4

$\sigma_{hp}{}^2$ component consists of variance due to h x p interaction; and $\sigma_{hps}{}^2$ due to h x s, p x s, and h x p x s interactions because the overall effects of period and season were precorrected.

TABLE 5.2. Relative values of estimated variance of sires effects using
different models and groupings

Model	Environmental grouping		
	A	B	C
	Period: 12 months	6 months	12 months
	Season: 4 months	2 months	1 month

Comparison of models within grouping

Model	A	B	C
1	1.00	1.00	1.00
2	0.84	0.83	0.63
3	0.68	0.60	0.51

Comparison of groupings within model

Model	A	B	C
1	1.00	1.14	1.36
2	1.00	1.13	1.00
3	1.00	0.99	1.00

All compared to model 1 using grouping A

Model	A	B	C
1	1.00	1.14	1.36
2	0.84	0.95	0.84
3	0.68	0.68	0.68

Model 1: herd-period-season fixed,
Model 2: herd-period fixed and herd-period-season random,
Model 3: herd fixed, herd-period and herd-period-season random.

TABLE 5.3. Ratios of the empirical to estimated variance of sire effects using different models and groupings

Model	Environmental grouping		
	A	B	C
	Period: 12 months Season: 4 months	6 months 2 months	12 months 1 month
1	1.14 (57)[†]	1.20 (72)	1.14 (79)
2	1.16 (44)	1.19 (55)	1.21 (44)
3	1.18 (36)	1.23 (37)	1.21 (37)

[†]Figures in parentheses are the averages of the empirical variances (kg^2)

TABLE 5.4. Correlations between sire effects for different models, and intraclass correlations among sire effects from all subsets using each model.


	Model 1	Model 2	Model 3

Grouping A: period of 12 months and season 4 months

Model 1	1.00		
Model 2	0.95	1.00	
Model 3	0.86	0.91	1.00

Grouping B: period of 6 months and season 2 months

Model 1	1.00		
Model 2	0.93	1.00	
Model 3	0.82	0.85	1.00

Grouping C: period of 12 months and season 1 month

Model 1	1.00		
Model 2	0.87	1.00	
Model 3	0.81	0.91	1.00

Intraclass correlations:

Grouping A	0.38	0.42	0.45
Grouping B	0.33	0.38	0.44
Grouping C	0.30	0.42	0.44

TABLE 5.5. Predicted effective number of daughters required using different
models to give the equivalent estimated variance of sire effects
for fat yield with different environmental groupings.

Desired estimated variance (kg^2) of sire effect	Model 1	Model 2	Model 3
Grouping A : Period of 12 months and season 4 months			
9	91.2	78.4	64.0
16	51.3	44.1	36.0
25	32.8	28.2	23.0
36	22.8	19.6	16.0
49	16.8	14.4	11.8
Grouping B : Period of 6 months and season 2 months			
9	90.6	76.6	55.4
16	51.0	43.1	31.2
25	32.6	27.6	20.0
36	22.7	19.2	13.9
49	16.6	14.1	10.2
Grouping C : Period of 12 months and season one month			
9	91.1	59.8	48.5
16	51.3	33.6	27.3
25	32.8	21.5	17.5
36	22.8	14.9	12.1
49	16.7	11.0	8.9

Fig 5.1. Empirical variances for each sire plotted for a model with herd-year (HY) fixed and herd-year-month (HYM) random effects against those for a model with herd-year-season fixed effects. The straight line has a slope of 1.00.
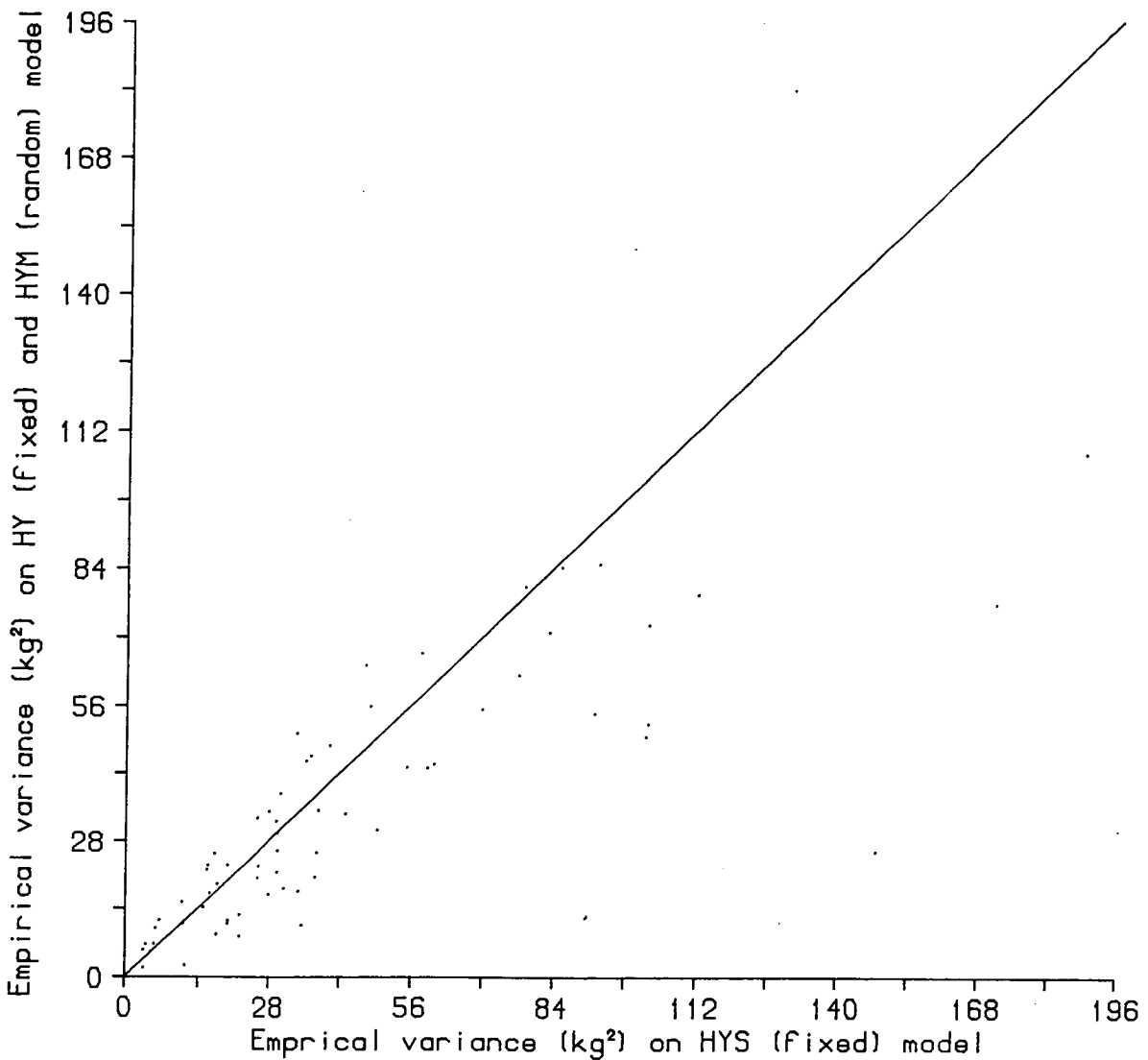
Chapter 6

ESTIMATING VARIANCES AND COVARIANCES OF MONTHS WITHIN HERD-YEAR FOR MILK PRODUCTION TRAITS

INTRODUCTION

In a herd-year-season model of sire evaluation, seasons are formed by grouping several consecutive months together. In a sire evaluation model fitting herd-year-season effects as fixed, it is assumed that the covariances between all cows calving in the same or different months are zero. When the season effects within herd-year are regarded as random and herd-year effects are regarded as fixed, it is assumed that all cows calving in different months but the same season have equal covariances with each other, and that the covariances of any of them with any of those calving in a different season are zero. However, in a model in which only herd effects are fixed and season within herd-year and year effects within herd are random, the covariances between cows calving in the same as well as different seasons are assumed to be non-zero.

However, since the intra-herd environmental variations from month to month (i.e. the herd-year x month interactions) are usually large (Chapter 3), it is not appropriate to assume that the covariances between all cows calving in different months of a season, are equal or are zero. The environmental correlations among milk production records in the same herd-year-season have been found to be slightly

smaller when seasons are longer (Chapter 4), showing that the environmental covariance between cows decreases as calving dates become further apart. Van Vleck (1966) reported the environmental correlations between milk yields of cows calving from zero to 18 year-seasons apart. No specific trends in the correlations were observed as the distance between year-seasons increased. However, his estimates were biased because they involved the covariances of each record with the rest in a herd, since each record was expressed as a deviation from the herd mean.

This study was undertaken to estimate the covariances between the first lactation records of cows calving in the same month and between records of those calving various months apart. It was assumed that the covariances of records between any pair of months the same distance apart are equal. For example, taking the recording year "December-November" currently used by the Milk Marketing Board of England and Wales (MMB) the one month apart covariance is the covariance of records of December with January, January with February, ..., October with November; and the two months apart: December with February, January with March, ..., September with November, and so on.

MATERIAL AND METHODS

The data consisted of the first lactation fat yield records of 8 recording years (December, 1972 to November, 1980) on 43089 progeny of 69 widely used proven Holstein-Friesian sires in 1533 herds participating in the Dairy Progeny Testing Scheme of the MMB, as described in Chapter 3. Records of widely used sires were used so that the environmental effects could be estimated more precisely.

70

Rather than analysing all data together, it was considered more appropriate to split them by herd at random into two subsets (22077 and 21012 records) such that all records of a herd were included in any one of the subsets, and compare the results from both. Each record was represented by the following model:

$$Y_{ijklm} = \mu + HY_{ij} + M_k + (HYxM)_{ijk} + Z_1 + bA + b'A^2 + e_{ijklm}$$

where:

Y_{ijklm} = record of the mth daughter of the lth sire calved at age A in the jth year and the kth month of the ith herd

μ = overall mean

HY_{ij} = fixed joint effect of the ith herd and the jth year of calving

M_k = fixed effect of the kth month of calving

$(HYxM)_{ijk}$ = random effect common to records of cows calved in the kth month of the jth year of the ith herd (random applies to Henderson's methods, 1953)

Z_1 = fixed effect of the lth sire (fixed effect since proven sires)

b and b' = linear and quadratic regression coefficients on age at calving A

e_{ijklm} = random error

This model was fitted to the data using the LSML76 computer program of Harvey (1977). Using the least squares constants from the above analysis, the data were adjusted for the effects of sire, month and age at calving. The herd-year effects were removed by subtracting the respective herd-year mean (calculated from records adjusted for sire, month and age) from each record. The herd-year mean included the record itself also. These deviated records were then used to calculate sums of squares (SSQ) for each month and sums of crossproducts (SCP) between different records in the same month and between different months. All calculations of SSQ and SCP were done on a herd-year basis, therefore the SSQ and SCP and their

coefficients (the numbers on which these SSQ and SCP were based) were accumulated, accordingly. Altogether, there were 12 SSQ and 12 SCP for the same month (i.e. 0-month-apart), and 66 SCP for different months. The figure 66 consists of 11 1-month-apart SCP and 10, 9, 8, 7, 6, 5, 4, 3, 2 and 1, respectively, 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10- and 11-months-apart SCP, the 11-months-apart figure being for between December and November. Since the herd-year mean has been subtracted from each record, these SSQ and SCP include covariances with other months, in addition to covariance between the months in question. Therefore, the equations of the expectation of these SSQ and SCP, in terms of the unknown variance and covariances, are given below.

For the same month:

$$E(SSQ) = n_i\, E(Y_{ij}-Y_{..})^2$$

$$= n_i V - 2n_i E(Y_{i.},Y_{..}) + n_i E(Y_{..})^2 \qquad [1]$$

$$E(SCP) = n_i(n_i-1)C_0 - 2n_i(n_i-1)E(Y_{i.},Y_{..}) + n_i(n_i-1)E(Y_{..})^2 \qquad [2]$$

For different months:

$$E(SCP) = E[(n_i n_j)(Y_{i.}- Y_{..})(Y_{j.}- Y_{..})] \qquad [3]$$

$$= n_i n_j[E(Y_{i.},Y_{j.}) - E(Y_{i.},Y_{..}) - E(Y_{j.},Y_{..}) + E(Y_{..})^2]$$

where:

$$E(Y_{i.},Y_{..}) = [\ n_i V + n_i(n_i-1)C_0 + \sum_{j \neq i} n_i n_j C_{|i-j|}\ ]/n_i N$$

$$E(Y_{i.},Y_{j.}) = C_{|i-j|}$$

$$E(Y_{..})^2 = [\ \sum_{i=1}^{P} n_i V + \sum_{i=1}^{P} n_i(n_i-1)C_0 + 2\sum_{i>j}^{P} n_i n_j C_{|i-j|}\ ]/N^2$$

E   = denotes the sign for expectation

Y_{ij} = the jth record in the ith month adjusted for the effects of sire, month and age of calving

$Y_{i.}$ = mean of the ith month within the herd-year (i=1,2,3,...,12)

$Y_{j.}$ = mean of the jth month within the herd-year (j=1,2,3,...,12)

$Y_{..}$ = herd-year mean

N = total number of records in the herd-year

P = total number of months (i.e. 12) within the herd-year

n_i = number of records in the ith month within the herd-year

n_j = number of records in the jth month within the herd-year

C_0 = 0-month-apart covariance i.e. the covariance between records

in the same month

V = residual variance within month plus C_0

$C_{|i-j|}$ = the $|i-j|$ months-apart covariance, where $i \neq j$;

and the sign $|\quad|$ denotes the absolute value

The equations for the expectations of the 90 terms of SSQ and SCP within a herd-year, in matrix notation, can be written as,

$$Xb = Y \qquad [4]$$

Where:

X = matrix (order 90x13) of the coefficients of unknown

parameters of variance and covariances,

b = vector (order 13x1) of unknown parameters of variance and

covariances, the first parameter being the V (as defined in

[1], [2] and [3]) followed by 0 to 11 months apart covariances,

Y = vector (order 90x1) of SSQ and SCP based on records deviated

from herd-year mean, the first 12 elements are the respective

SSQ for different months followed by 11 elements for 0-month-apart

SCP, 10 for 1-month-apart, ..., 1 for 11-months-apart SCP.

It was observed that all rows of the matrix X sum to zero but neither the columns of matrix X nor the elements of vector Y sum to zero. According to [1], [2] and [3], altogether there are 90 equations in [4] for the expectations of SSQ and SCP. The matrix X

and the vector Y can be transformed to reduce the total number of equations to a number equal to the number of unknown parameters in the vector b. All equations derived from [1] can be added into a single equation and all those from [2] into a second equation. Likewise the 66 equations derived from [3] are reduced to only 11 equations since the equations pertaining to the expectations of the same months apart SCP are added together. These 11 equations need to be multiplied by two in order to make the transformed matrix X (i.e. X^*) symmetric. This transformed matrix X^* is now of order 13x13 and elements of each row and each column sum to zero. Similarly, elements of the transformed vector Y^* (order 13x1) also sum to zero. The equations pertaining to 1, 2, 3, ..., 11 months apart unknown covariances are multiplied by two since the expectations of the between months SCP of only one triangle of the matrix of SCP were calculated in order to minimize the computations (see Table 6.1 and calculations in the example given in the appendix at the end of this Chapter). In view of the above characteristics of the matrix X the solution to unknown parameters of variance and covariances in the vector b can be obtained by either one of the following two methods:

(a) using untransformed matrix X and vector Y,

$$X'Xb = X'Y$$

$$b = (X'X)^{-1}X'Y \qquad \qquad [5]$$

(b) using transformed matrix X^* and vector Y^*,

$$X^*b = Y^*$$

$$b = (X^*)^{-1}Y^* \qquad \qquad [6]$$

The premultification of both sides of equation [4] by the transpose of matrix X, forces the matrix X'X to be symmetric (order 13x13). The vector X'Y is eventually of order 13x1. Although the matrices

X'X and X* are symmetric, they are not of full rank, and hence there is no unique solution for the unknown parameters in the vector b. Because of using deviations within herd-year all covariances can not be estimated. Therefore, in order to solve these equations, it was considered reasonable to impose a constraint that the longest distance apart covariance (i.e. the 11-months-apart within herd-year analyses) is zero. This caused all elements of the 13th row and the 13th column of the matrices X'X and X* to be zero, and also the 13th elements of the vectors X'Y and Y* were automatically zero. As a consequence, the matrices X'X and X* (both having order 12x12) were symmetric matrices of full rank (i.e. 12). Solution of the unknown parameters in the vector b can then be obtained by inverting the matrices X'X or X* and premultiplying the inverse by the corresponding right hand side vector.

The covariances between year-months within herd-period (period of 24 months) and between year-months within herd (period of 96 months, i.e. the whole data available in a herd) were also estimated using the method (b). Method (b) requires a much smaller store for the transformed matrix X* than does the untransformed matrix X in method (a).

The variances and covariances within and between months, using records deviated from herd-year mean, were also calculated by dividing each SSQ and SCP by the corresponding coefficients (i.e. the numbers on which these SSQ and SCP were based). The variances of records within different months were not much different from each other. But the covariances between months at the same distance (e.g. 0-month-apart: for December, January, ..., November; 1-month-apart: between December and January, January and February,

..., October and November; 2-months-apart: between December and February, January and March, ..., September and November) were found to be very heterogenous (Table 6.1). These variances and covariances, obtained from deviated records, were broken up in terms of the unknown variance and covariances in order to examine the bias in them caused by subtracting the herd-year mean from each record. For these investigations, the coefficients of the unknown variances and covariances in the matrix X were scaled by the corresponding coefficients (numbers) on which the SSQ and SCP were based (see example in the appendix). This procedure modifies the matrix X as if each SSQ, or each of the same months-apart SCP, are based on equal coefficients (i.e. as if the numbers of records in each month are equal). The scaling procedure in matrix notation is given below. As far as the estimation of the unknown variance and covariances is concerned, the properties of the scaled equations remain unchanged.

$$D^{-1}Xb = D^{-1}Y \qquad\qquad [7]$$

where:

D = diagonal matrix (order 90x90) having the coefficients of SSQ and SCP on its diagonal corresponding to the respective equation of the expectation.

The matrix X and the vectors b and Y are the same as defined in [4].

Most of the calculations for estimating the unknown variance and covariances using method (b) are illustrated in the appendix (at the end of this chapter) using an example.

It could be asked what is the relationship between the estimated value of C_0 (i.e. the 0-month-apart covariance) and the component of

variance for month (σ_m^2) within herd-year, given in Chapter 4. This question was answered using the following formula, which gives an approximate value for the month component from the estimated values of different months apart covariances (R. Thompson, 1985, personal communication).

$$\sigma_m^2 = [\ PC_0 - (PC_0 + 2 \sum_{i=1}^{P-1} (P-i)C_i)/P]/(P-1) \qquad\qquad [8]$$

where:

P = number of months within a period (period of 6 or 12 months) minus one

C_i = i-months-apart covariance

RESULTS AND DISCUSSION

Since the inferences from the results of both subsets of data were quite similar, the full results from only one subset (22077 records) are presented. However, the estimates of various months-apart covariances from within herd-year analyses for both subsets are plotted in Figure 6.2.

The within-month variances and the covariances between months based on records deviated from herd-year mean are given in Table 6.1. These estimates are biased because they include the covariances with other months. This bias was introduced by the subtraction of the herd-year mean from each record. In order to demonstrate the consequences of subtracting the herd-year mean from each record, the composition of only the 0-month-apart scaled SCP in terms of the unknown variance and covariances is shown in Figure 6.1. Only the scaled coefficients, obtained from [7], of unknown variance and 0- and 1-month-apart covariances from the equations of the expectations of 0-month-apart SCP have been plotted (Fig 6.1).

The scaled coefficients of other unknown covariances (i.e. for 2 to 11 months apart covariances) in these equations were close to zero, indicating that their contribution to the bias in 0-month-apart covariances were small. From Figure 6.1, it was clear that there were unequal contributions from the unknown variance and covariance terms to each of the 0-month-apart covariances shown in Table 6.1 (based on deviated records). These contributions, however, did not show any specific trend.

The estimated values of 0 to 11-months-apart covariances obtained from the solution of the equations in [5] and [6] for within herd-year analyses have been plotted in Figure 6.2. The trends in the estimates of covariances from both subsets were quite similar. The solutions from both methods, [5] and [6], were not identical as different weightings are accounted for in these two procedures, but the estimates showed quite similar tends. The covariances from 0 through to 5-months-apart showed an almost linear decline, and thereafter all covariances were apparently similar and close to zero. The variation in the estimates of 6 to 10-months-apart covariances (the 11-months-apart covariance was forced to be zero) can be attributed to the fact that, as compared to the 0- to 5-months-apart covariances, they were based on smaller coefficients, e.g. the 10-months-apart covariance comprised the covariance between records in December and October, and January and November; and the 11-months-apart covariance was only between December and November.

The estimates of covariances from the analyses over 24 months and 96 months also showed a linear decline from 0 through to 5 or 6 months apart (Figures 6.3 and 6.4). All covariances beyond 6 months apart showed a gradual declining trend, although very slow,

suggesting periodic trends in the covariances between year-months. In the absence of periodic trends all covariances beyond 11 months apart should have been similar, having no specific trends.

Since there were no unique solutions to the parameters of variance and covariance in [5] or [6] the longest distance-apart-covariances in all analyses were assumed to be equal to zero. Therefore, there does not seem to be an exact correspondence between the estimates from different analyses, as the estimates of the covariances from the analyses within 24 months or 96 months include the between years within herd-period covariances in addition to the covariances between the corresponding year-months. However, the following two checks were made to test the consistency of the estimates of the covariances. Firstly, the within-month variances from all analyses, using the same data subset (22077 records), were estimated as the value of the parameter V minus C_0, and are given below:

Analyses	within month variance (kg^2)
within herd-year, using method (a)	687
within herd-year, using method (b)	672
within herd-period (period of 24 months), using method (b)	657
within herd-period (period of 96 months), using method (b)	651

These variances are close to the residual variance in the herd-year-month analysis (i.e. 690 kg^2), given in Chapter 4, showing that the estimates of the covariances from different period groupings are consistent. Secondly, using the estimated values of various months apart covariances from within herd-year analysis, the

components of variance for month within herd-period with a period of 6 or 12 months were estimated to be 64 and 103 kg^2 using equation [8]. These estimates are in agreement with the values of 49 and 85 kg^2 respectively obtained with periods of either 6 or 12 months, in Chapter 4.

The main conclusion from the results of this study is that the differences in the estimates of covariances from 0 to 5-6 months apart were quite substantial and thereafter all covariances were observed to be similar. Therefore, a model assuming realistic variance-covariances between months may be useful to improve progeny group comparisons. If it is necessary to fit herd-year effects as fixed then, in view of the observed covariances and the trends in them, the following two models of sire evaluation may be suitable:

(i) taking account of the variance-covariances between all months within a herd-year,

(ii) assuming a linear decline in the covariances from 0 through to 5 months apart and all covariances beyond 5 months apart to be zero.

The latter model may be helpful from the point of computations as the covariances between several pairs of months can be taken to be zero, a priori.

The year effects within herd are mainly fitted to account for the environmental trends within herds. However, from the plot of 0 to 95 months apart covariances in Figure 6.4 it was observed that yearly trends are apparently small. Therefore, year effects within herd could be replaced by a period effect (e.g. period of 24 months). This consideration may not be of any help in a herd-period-year-season fixed effects model, but it might be useful

in a model in which herd-period effects are fitted as fixed and herd-period-year-season effects as random. This is because it would increase the number of comparisons from the herd-year subclasses having all or almost all records of the same sire. Similarly, a period effect instead of the year effects may also be helpful in a model in which covariances between year-months within herd-period are accounted for.

SUMMARY

A procedure for estimating the covariances among first lactation records (precorrected for sire, age and month of calving, and deviated from the respective herd-year mean) commencing in the same and different months within herd-year has been given. Covariances were estimated to examine an assumption underlying the commonly used herd-year-season fixed effects model of sire evaluation; the assumption being that all covariances between cows calving in the same or different months are zero. Analyses were conducted on two independent subsets of data comprising 22077 and 21012 first lactation fat yield records on progeny of 69 widely used Holstein-Friesian sires in 1533 herds. The covariances from 0 through to 5 months apart were observed to decline in an almost linear fashion as the distance between months (within herd-year) increased. All covariances beyond 5 months apart were small. It was concluded that a sire evaluation model accounting for these covariances might be desirable.

Appendix to Chapter 6.

Example using herd-year data over 4 months only with a
maximum of 5 records in a month. These data are assumed
to be precorrected for all effects except for the herd-year
effect.

Total number of records = 13
Herd-year mean   = 20.19

The records deviated from herd-year mean:

Month	Individual records deviated from herd-year mean					Total no. in month
1	-4.19	4.81	2.81	1.21	0.61	5
2	-0.39	2.01	-2.89			3
3	-3.29	0.51	-2.19	3.61		4
4	-2.59					1

Sums of squares (SSQ) and sums of crossproducts (SCP):

	Month			
	1	2	3	4

SSQ for different months:

	1	2	3	4
SSQ:	50.40	12.54	28.92	6.72
Coeff.:	5	3	4	1

SCP (between different records in the same month):

	1	2	3	4
SCP	-22.96	-10.92	-27.04	0.00
Coeff.:	20	6	12	0

SCP (between months)[†] :

	1	2	3	4
month-1	....	15	20	5
month-2	-6.69	....	12	3
month-3	-7.17	1.75	....	4
month-4	-13.58	3.31	3.55	....

[†]figures in the lower triangle are the SCP between
months and in the upper triangle the corresponding
coefficients on which SCP are based.

The equations for the expectations of the SSQ and SCP
using equations [1], [2] and [3]:

	V	c_0	c_1	c_2	c_3
E(Y..,Y..) =	0.0769	0.2249	0.3669	0.2722	0.0592
E(Y1.,Y..) =	0.0769	0.3077	0.2308	0.3077	0.0769
E(Y2.,Y..) =	0.0769	0.1538	0.6923	0.0769	0.0000
E(Y3.,Y..) =	0.0769	0.2308	0.3077	0.3846	0.0000
E(Y4.,Y..) =	0.0769	0.0000	0.3077	0.2308	0.3846
E(50.40) =	4.6154	-1.9527	-0.4734	-1.7160	-0.4734
E(12.54) =	2.7692	-0.2485	-3.0533	0.3550	0.1775
E(28.92) =	3.6923	-0.9467	-0.9941	-1.9882	0.2367
E(6.72) =	0.9231	0.2249	-0.2485	-0.1893	-0.7101
E(-22.96) =	-1.5385	12.1893	-1.8935	-6.8639	-1.8935
E(-10.92) =	-0.4615	5.5030	-6.1065	0.7101	0.3550
E(-27.04) =	-0.9231	9.1598	-2.9822	-5.9645	0.7101
E(-6.69) =	-1.1538	-3.5503	6.6568	-1.6864	-0.2663
E(1.75) =	-0.9231	-1.9172	4.4024	-2.2722	0.7101
E(3.55) =	-0.3077	-0.0237	3.0059	-1.3728	-1.3018
E(-7.17) =	-1.5385	-6.2722	-3.4320	11.5976	-0.3550
E(3.31) =	-0.2308	0.2130	-1.8994	2.8935	-0.9763
E(-13.58) =	-0.3846	-0.4142	-0.8580	-1.3314	2.9882

C = unknown covariance among records, the subscript followed
    by C indicate the "months-apart";
V = the variance within month plus c_0.

The equations for the expectations of SSQ and SCP can be
written in matrx notation as below:

$$Xb = Y$$

X = matrix of the coefficients of unknown variance and covariances
    in the equations of the expectations of SSQ and SCP,

Y = vector of SSQ and SCP, based on deviated records,

b = vector of parameters of unknown variance and covariances.

$$
X = \begin{bmatrix}
4.6154 & -1.9527 & -0.4734 & -1.7160 & -0.4734 \\
2.7692 & -0.2485 & -3.0533 & 0.3550 & 0.1775 \\
3.6923 & -0.9467 & -0.9941 & -1.9882 & 0.2367 \\
0.9231 & 0.2249 & -0.2485 & -0.1893 & -0.7101 \\
-1.5385 & 12.1893 & -1.8935 & -6.8639 & -1.8935 \\
-0.4615 & 5.5030 & -6.1065 & 0.7101 & 0.3550 \\
-0.9231 & 9.1598 & -2.9822 & -5.9645 & 0.7101 \\
-1.1538 & -3.5503 & 6.6568 & -1.6864 & -0.2663 \\
-0.9231 & -1.9172 & 4.4024 & -2.2722 & 0.7101 \\
-0.3077 & -0.0237 & 3.0059 & -1.3728 & -1.3018 \\
-1.5385 & -6.2722 & -3.4320 & 11.5976 & -0.3550 \\
-0.2308 & 0.2130 & -1.8994 & 2.8935 & -0.9763 \\
-0.3846 & -0.4142 & -0.8580 & -1.3314 & 2.9882
\end{bmatrix} ; Y = \begin{bmatrix}
50.40 \\
12.54 \\
28.92 \\
6.72 \\
-22.96 \\
-10.92 \\
-27.04 \\
-6.69 \\
1.75 \\
3.55 \\
-7.17 \\
3.31 \\
-13.58
\end{bmatrix}
$$

The transformed matrix X^* and vector Y^* are,

$$X^* = \begin{bmatrix} 12.0000 & -2.9231 & -4.7692 & -3.5385 & -0.7692 \\ -2.9231 & 26.8521 & -10.9822 & -12.1183 & -0.8284 \\ -4.7692 & -10.9822 & 28.1302 & -10.6627 & -1.7160 \\ -3.5385 & -12.1183 & -10.6627 & 28.9822 & -2.6627 \\ -0.7692 & -0.8284 & -1.7160 & -2.6627 & 5.9763 \end{bmatrix} \; ; \; Y^* = \begin{bmatrix} 98.59 \\ -60.92 \\ -2.79 \\ -7.72 \\ -27.16 \end{bmatrix}$$

Solution by inverting the transformed matrix X^* and vector Y^* and imposing a constraint that $C_3 = 0$, ,

$$b = \begin{bmatrix} V \\ C_0 \\ C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} 11.43 \\ 2.25 \\ 4.07 \\ 3.57 \\ 0.00 \end{bmatrix}$$

The procedure for scaling the coefficients of unknown variance and covariances in the equations of the expectations of SSQ and SCP.

$D^{-1}Xb = D^{-1}Y$

D = diagonal matrix (order 13x13) having coeff. of SSQ and SCP on its diag.

$$D = \begin{bmatrix} 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 15 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 20 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Scaled coefficients (X_S):

$$X_S = D^{-1}X = \begin{bmatrix} 0.9231 & -0.3905 & -0.0947 & -0.3432 & -0.0947 \\ 0.9231 & -0.0828 & -1.0178 & 0.1183 & 0.0592 \\ 0.9231 & -0.2367 & -0.2485 & -0.4970 & 0.0592 \\ 0.9231 & 0.2249 & -0.2485 & -0.1893 & -0.7101 \\ -0.0769 & 0.6095 & -0.0947 & -0.3432 & -0.0947 \\ -0.0769 & 0.9172 & -1.0177 & 0.1183 & 0.0592 \\ -0.0769 & 0.7633 & -0.2485 & -0.4970 & 0.0592 \\ -0.0769 & -0.2367 & 0.4438 & -0.1124 & -0.0178 \\ -0.0769 & -0.1598 & 0.3669 & -0.1893 & 0.0592 \\ -0.0769 & -0.0059 & 0.7515 & -0.3432 & -0.3255 \\ -0.0769 & -0.3136 & -0.1716 & 0.5799 & -0.0177 \\ -0.0769 & 0.0710 & -0.6331 & 0.9645 & -0.3254 \\ -0.0769 & -0.0828 & -0.1716 & -0.2663 & 0.5976 \end{bmatrix}$$

84

TABLE 6.1. Variances within month and covariances between months based
on records of fat yield (kg) deviated from herd-year mean
after adjustments for sire, age and month of calving.

	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV
V	699	645	591	555	605	586	672	717	640	655	695	614
N(V)	1565	1403	1449	1343	960	670	612	1235	2390	4405	3480	2565
C_0	20	67	21	18	7	26	31	21	-14	-12	-19	-6
N(C_0)	4266	3446	4182	4922	2250	1338	1664	8838	10546	25806	15502	8130
DEC[†]	...	2217	1571	1452	749	468	576	1153	1790	3655	3905	2957
JAN	0	...	1987	1508	763	468	422	1350	1576	3075	2802	1665
FEB	9	14	...	2467	847	505	467	2697	2086	2868	2383	1876
MAR	11	14	31	...	1525	434	666	2182	1370	2840	1805	1472
APR	-31	-93	-50	-4	...	981	620	715	964	1950	1475	1102
MAY	-59	-55	-41	-33	-5	...	546	459	564	951	930	715
JUN	-56	-75	-45	23	12	-20	...	1035	641	969	773	767
JUL	-58	-153	-77	-83	26	-52	-101	...	2658	2579	2102	1334
AUG	-74	-84	-70	-94	-120	-120	-76	-20	...	8307	3505	2704
SEP	-99	-87	-72	-90	-100	-97	-137	-53	-24	...	11563	5848
OCT	-70	-94	-86	-122	-59	-82	-144	-34	-55	-40	...	6706
NOV	-96	-110	-114	-107	-68	-84	11	-27	-52	-40	-20	...

V = residual variance i.e. variance within month,
C_0 = Covariance between different records in the same month,

N = the coefficients on which V and C_0 are based,

[†] Figures in the lower triangle are the covariances between different
months, and in the upper triangle the corresponding coefficients.

Fig. 6.1. Composition of 0-month-apart covariances, based on records deviated from herd-year mean, in terms of the unknown variance and covariances (22077 records).
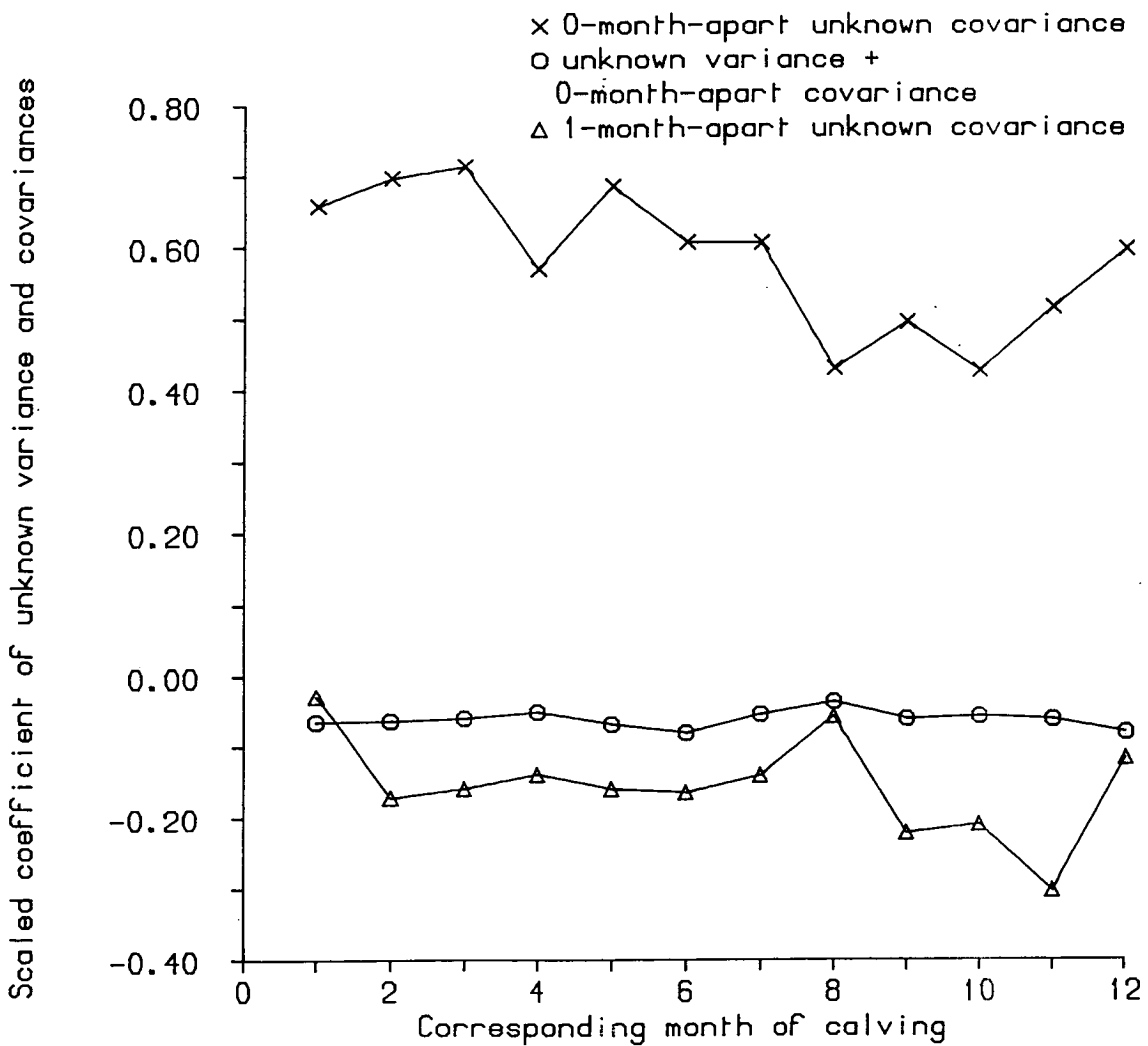
Fig. 6.2. Various months apart covariances
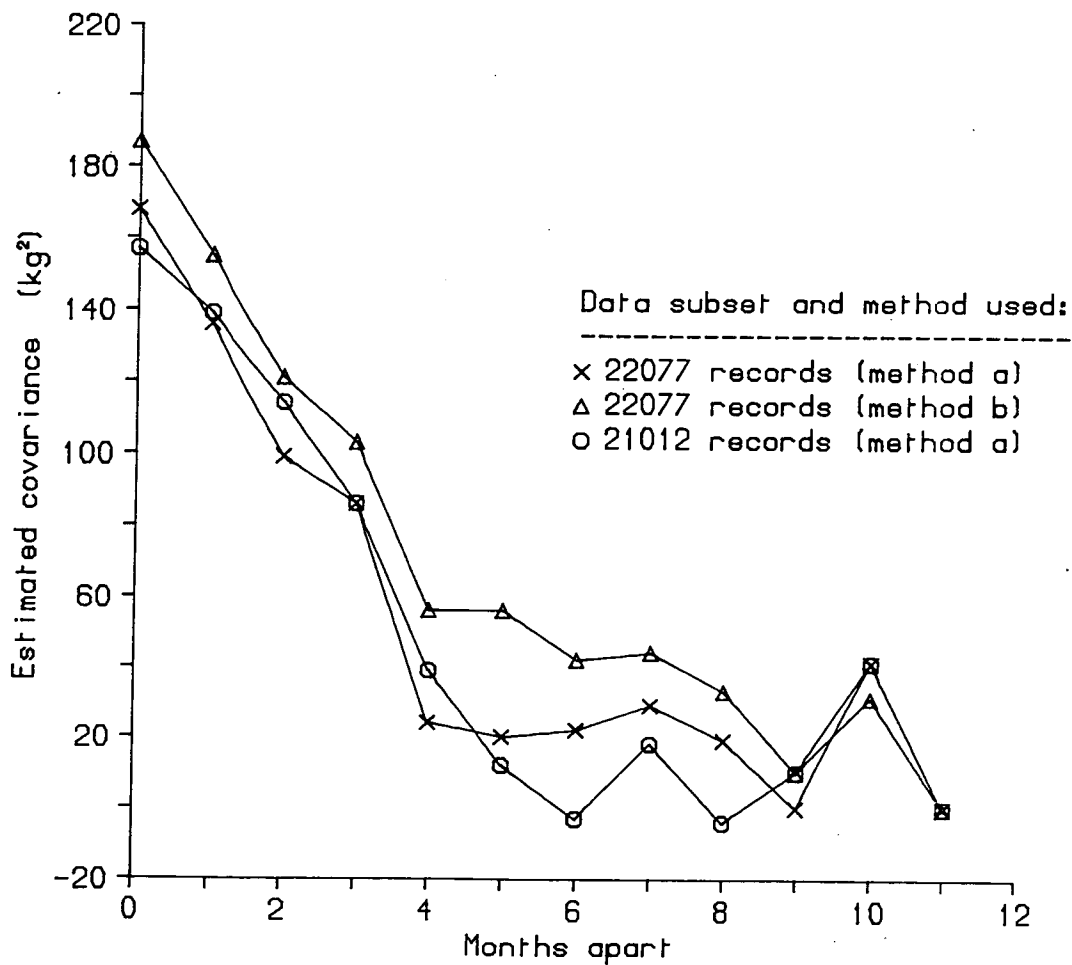among fat yield records within
herd-year

Data subset and method used:

× 22077 records (method a)
△ 22077 records (method b)
o 21012 records (method a)

Fig. 6.3. Various months apart covariances among
fat yield records within herd-period
(using a period of 24 months, 22077
records and method b).
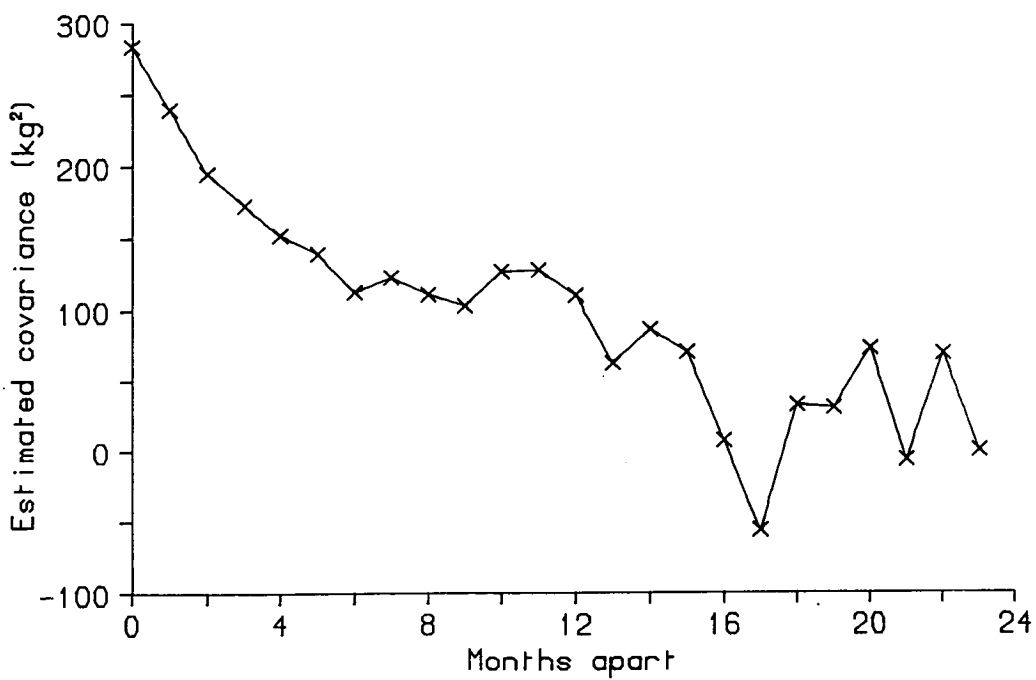
Fig. 6.4. Various months apart covariances among fat yield records within herd using all data over 96 months (using 22077 records and method b).
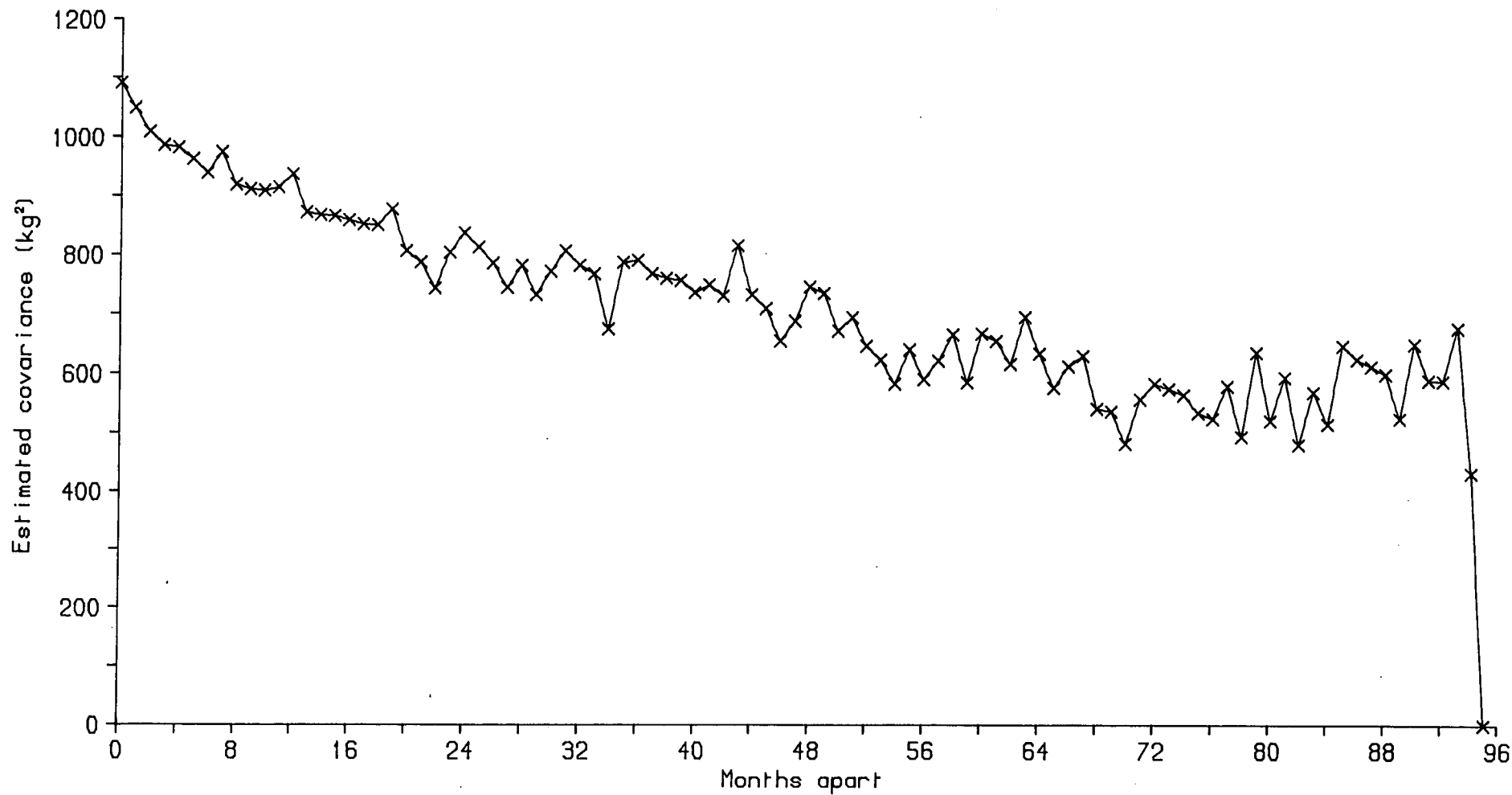
Chapter 7

SIRE EVALUATION FITTING A LINEAR COVARIANCE STRUCTURE BETWEEN MONTHS WITHIN HERD-YEAR ("ROLLING MONTHS" MODEL)

INTRODUCTION

In sire evaluation models used in most breeding schemes it is assumed that the covariances between cows calving in the same and different months in the same herd-year are zero, since herd-year-seasons are fitted as fixed effects. However, the covariances between months in the same herd-year have been observed to decline from 0 through to 5 months apart in an almost linear fashion as the distance between months increased (Chapter 6). All covariances beyond 5 months apart were small.

The aim of this study is to derive a sire evaluation model which could take account of these observed covariances between months. In deriving this model it will be assumed that the covariances between months decline linearly as the distance between them increases, in order to mimic the observed trends. Models such as this have been recently suggested for the analyses of crop variety trials (Bartlett, 1978; Wilkinson, Eckert, Hancock and Mayo, 1983).

MATERIAL AND METHODS

Data:

The same 8 subsets of data as described in Chapter 5 were used for estimating sire effects. These data were the first lactation fat yield records, precorrected for the effect of month of calving, of 49242 progeny of 69 widely used proven Holstein-Friesian sires in herds participating in the Dairy Progeny Testing Scheme (DPTS) of the Milk Marketing Board of England and Wales (MMB). The data on widely used sires were useful for empirical comparisons of sire effects from different subsets.

Methods:

The observed covariances between months could have been taken into account using a procedure similar to the sires's relationships matrix (Henderson, 1975). But in this study it has been chosen to derive a model in which it is assumed that the covariances between months decline linearly as the distance between them increases. An assumption, that the covariances between months decrease linearly from 0 through to 5 months apart, is close to the observed relationship between the covariance and the distance between months (see Figure 7.0). However, in order to investigate the appropriate model alternative analyses will be conducted assuming that the covariances between months decline linearly from 0 through to either 1- or 2- or 3- or 4- or 5- or 8-months-apart and the rest of the between months covariances are zero. This model hereafter will be called the "rolling months" model as it accounts for the covariances of each month with a "specified" number of months on each side of

91

it. Suppose the covariances between months are assumed to decline linearly from 0 through to W months apart, then the covariances among '2W-1' months are accounted for with each individual month — with the exception of those of the beginning or of the end of the herd-year blocks. The covariances of each month with its preceding as well as with the following months are assumed to decline linearly. The value of W will hereafter be referred to as the "width" of a group of months among which the covariances are accounted for. In order to obtain the assumed variance-covariance structure between months a "pseudo-month" effect will be fitted. The total number of pseudo-month effects to be fitted within herd-year will be 'm+W-1' where 'm' is the number of months.

The records were assumed to be represented by the following model:

$$Y = Pb_1 + Hb_2 + Cb_3 + Zu + e \qquad [1]$$

where,

Y = vector of the first lactation fat yield records,

P = design matrix for pseudo-months resulting from the assumed variance-covariance structure between months, with variance of month effects equal to $W\phi$ and covariance of month effects in the same herd-year, i-months-apart equal to $(W-i)\phi$ for $1 \leqslant i < W$ and equal to 0 for $i \geqslant W$

H = design matrix for herd-years,

C = matrix of information on covariables (ie age and square of age at calving),

Z = design matrix for sires,

b_1 = vector of unknown random effects for pseudo-months,

b_2 = vector of unknown fixed effects for herd-years,

b_3 = vector of unknown linear and quadratic regression coefficients of

fat yield on age at calving,

u = vector of unknown fixed effects for sires, (fixed since proven sires),

λ = ratio of residual variance to the variance of records within the same pseudo-month, σ_e^2/ϕ

e = vector of residuals

The herd-year effects have been chosen to fit in the above model (i.e. the rolling months model) in order to compare it with the "random month" model in Chapter 5, in which herd-year effects were fitted as fixed and instead of a pseudo-month effect the random effect of herd-year-month was fitted with the rest of the effects the same as in the rolling months model. Alternatively, a herd-period effect (period of 2 years) could have been fitted instead of herd-year effects, as the yearly trends in the covariances between year-months over two consecutive years were small (Chapter 6). A rolling months model, in which the covariances among records only in the same month are accounted for and the rest of the covariances between months are assumed to be zero (i.e. W=1), is equivalent to the random month model. However, the rolling months model, with an appropriate width, has the advantage over the random month model or a herd-year-season fixed effects model, in that because it gives different weightings to the comparisons at different times apart.

The mixed model equations (MME) pertaining to the rolling months model are given below in matrix notation.

$$
\begin{bmatrix}
P'P+\lambda I & P'H & P'C & P'Z \\
H'P & H'H & H'C & H'Z \\
C'P & C'H & C'C & C'Z \\
Z'Z & Z'P & Z'C & Z'Z
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3 \\
u
\end{bmatrix}
=
\begin{bmatrix}
P'Y \\
H'Y \\
C'Y \\
Z'Y
\end{bmatrix}
\qquad [2]
$$

All the submatrices in [2], except those in the equations for pseudo-months, can be constructed by the usual procedure. Assuming that the data are available over only 4 months in a herd-year and the width (W) for rolling over months is 3, an example is given below for constructing the P'P and P'H submatrices. The submatrices P'C, P'Z and P'Y can be constructed similarly. Suppose n_1, n_2, n_3 and n_4 are the numbers of records respectively for months m_1, m_2, m_3 and m_4, then the total number of pseudo-months (P) effects to be estimated will be 6 (i.e. m+W-1 within a herd-year).

Pseudo-month

	1	2	3	4	5	6
1	$n_1+\lambda$	n_1	n_1	0	0	0
2	n_1	$n_1+n_2+\lambda$	n_1+n_2	n_2	0	0
3	n_1	n_1+n_2	$n_1+n_2+n_3+\lambda$	n_2+n_3	n_3	0
4	0	n_2	n_2+n_3	$n_2+n_3+n_4+\lambda$	n_3+n_4	n_4
5	0	0	n_3	n_3+n_4	$n_3+n_4+\lambda$	n_4
6	0	0	0	n_4	n_4	$n_4+\lambda$

$$P'P = \text{(above matrix)}$$

$$P'H = \begin{bmatrix} n_1 \\ n_1+n_2 \\ n_1+n_2+n_3 \\ n_2+n_3+n_4 \\ n_3+n_4 \\ n_4 \end{bmatrix}$$

The month effects in terms of the pseudo-month effects can be given by the following equations.

$$m_i = P_i+P_{i+1}+\ldots+P_{i+W-1}$$

$$m_j = P_j+P_{j+1}+\ldots+P_{j+W-1}$$

The following assumptions are made in fitting the rolling months model:

(i) homogeneity of variance over pseudo-months (Var[P] = ϕ)

(ii) Cov $(P_i, P_j) = 0$; $i \neq j$

In order to fit the rolling months model the pseudo-month effects must be assumed to be random, otherwise the estimates of sire effects and their variances would be identical to those on a herd-year-month fixed effects model. Using the above assumptions the linear relationship between the covariance (Cov) and different months-apart distances (shown in Figure 7.0) can be shown or re-demonstrated as follows, using the same example with information over only 4 months and assuming W equal to 3,

$$\text{Cov}(m_1, m_1) = C_0 = \text{Cov}(P_1 + P_2 + P_3, P_1 + P_2 + P_3) = 3\phi.$$

$$\text{Cov}(m_1, m_2) = C_1 = \text{Cov}(P_1 + P_2 + P_3, P_2 + P_3 + P_4) = 2\phi$$

$$\text{Cov}(m_1, m_3) = C_2 = \text{Cov}(P_1 + P_2 + P_3, P_3 + P_4 + P_5) = \phi$$

$$\text{Cov}(m_1, m_4) = C_3 = \text{Cov}(P_1 + P_2 + P_3, P_4 + P_5 + P_6) = 0$$

where:

$\quad\quad C_i$ = i-months-apart covariance

$\quad\quad \phi$ = Variance of records in the same pseudo-month

The quantity ϕ was estimated using the covariance among records in the same month ($C_0 = 168 \text{ kg}^2$) given in Chapter 6.

Since the submatrix $P'P + \lambda I$ is not a diagonal matrix the equations for the pseudo-month effects may be absorbed either by inverting it or by successively eliminating the pseudo-month effects. In this study the latter procedure was used as it required less computing time than the inversion. The MME in [2] after

absorption of the equations for the effects of pseudo-month and herd-year can be written as below,

let:

$$S = [I - P(P'P+\lambda I)^{-1}P']$$

$$Q = [S - SH(H'SH)^{-1}H'S]$$

then the equations for age and sire effects would be,

$$\begin{bmatrix} C'QC & C'QZ \\ Z'QC & Z'QZ \end{bmatrix} \begin{bmatrix} b_3 \\ u \end{bmatrix} = \begin{bmatrix} C'QY \\ Z'QY \end{bmatrix} \qquad [3]$$

The estimates of sire effects were obtained by absorbing the equations for age into sire equations and then inverting the sire's coefficient matrix,

letting:

$$R = [Q - QC(C'QC)^{-1}C'Q]$$

then

$$u = (Z'RZ)^{-1}(Z'RY)$$

The residual mean squares (σ_e^2) were calculated by the procedure given by Thompson (1969).

$$\sigma_e^2 = [\sum_{i=1}^{h} Y'QY - Y'QC(C'QC)^{-1}C'QY$$
$$- Y'RZ(Z'RZ)^{-1}Z'RY]/[N - rank(fixed\ effects)] \qquad [4]$$

where, h = total number of herd-years
N = total number of records

$$V(u) = (Z'RZ)^{-1}\sigma_e^2 \qquad [5]$$

The MME for the random month model would be similar to [2] and [3] of the rolling months model. These two models differ only in assuming the variances and covariances between months. In the random month model all covariances between months are assumed to be zero, consequently, the $P'P+\lambda I$ submatrix in [2] for the month equations in the random month model would be diagonal. Therefore,

in this study only these two models will be compared with each other. Let the random month model be model 1 and the rolling months model be model 2.

Criteria of comparison:

The following two criteria were used to compare the rolling months model with the random month model:

(a) Empirical variance: the following formula given in Chapter 5 was used to calculate the empirical variances for each sire.

$$\text{Empirical variance} = [\sum_{j=1}^{P} (u_{ij} - \bar{u}_i)^2]/(P-1) \qquad [6]$$

where:

$$\bar{u}_i = (\sum_{j=1}^{P} u_{ij})/P$$

u_{ij} = sire effect for the ith sire from the jth subset of data,

P = number of subsets of data

(b) Comparing the predicted variances of the estimates of sire effects on the random month model given that the rolling months model is the true model (see Henderson, 1975a). The variance-covariance (V) of elements in the vector Y in [1] for both models can be written as follows:

$V_1 = I\sigma_e^2 + MM'\sigma_m^2$ ... (random month model)

$V_2 = I\sigma_e^2 + PP'\phi$ ... (rolling months model)

where: M = design matrix for herd-year-months

σ_m^2 = variance of records in the same month

Let the variance-covariance matrices of month effects for models 1 and 2 be denoted by A_1 and A_2, respectively. Then for a value of W equal to 5 months the matrices A_1 and A_2 can be written as follows:

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \sigma_m^2/\sigma_e^2 \qquad [7]$$

$$A_2 = \begin{bmatrix} 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 4 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 5 & 4 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 5 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 3 & 4 & 5 \end{bmatrix} \phi/\sigma_e^2 \qquad [8]$$

Then

$$V_1 = I + MA_1M'$$

$$V_2 = I + MA_2M'$$

and

$$V_2 - V_1 = M(A_2 - A_1)M'$$

Based on the same design matrix the predicted variances of the estimates of sire effects on the random month (model 1) and the rolling months (model 2) models can be written as,

let:

$$(X'V^{-1}X) = (Z'RZ)$$

then:

$$Var(u_1|model_1) = (X'V_1^{-1}X)^{-1}\sigma_{e1}^2 \qquad [9]$$

$$\text{Var}(u_2 \mid \text{model}_2) = (X'V_2^{-1}X)^{-1}\sigma_{e2}^2 \qquad [10]$$

$$\text{Var}(u_1 \mid \text{model}_2) = [(X'V_1^{-1}X)^{-1}X'V_1^{-1}V_2V_1^{-1}X(X'V_1^{-1}X)^{-1}]\sigma_{e2}^2 . \qquad [11]$$

When $V_2 = V_1$, [11] reduces to [9]. Given that model 2 (the rolling months model) is the true model, the efficiency of model 1 (the random month model) can be judged by comparing $\text{Var}(u_2 \mid \text{model}_2)$ and $\text{Var}(u_1 \mid \text{model}_2)$. The larger the value of $\text{Var}(u_1 \mid \text{model}_2)$ in relation to $\text{Var}(u_2 \mid \text{model}_2)$ the more efficient the rolling months model.

Computing procedure for Var(u₁ | model₂):

Equation [11] can be rewritten as follows:

$$\text{Var}(u_1 \mid \text{model}_2) = [(X'V_1^{-1}X)^{-1} +$$
$$(X'V_1^{-1}X)^{-1}X'V_1^{-1}(V_2-V_1)V_1^{-1}X(X'V_1^{-1}X)^{-1}]\sigma_{e2}^2 \qquad [12]$$

Replacing V_2-V_1 by $M(A_2-A_1)M'$:

$$\text{Var}(u_1 \mid \text{model}_2) = [(X'V_1^{-1}X)^{-1} +$$
$$(X'V_1^{-1}X)^{-1}(X'V_1^{-1}M(A_2-A_1)M'V_1^{-1}X(X'V_1^{-1}X)^{-1}]\sigma_{e2}^2 \qquad [13]$$

In order to compute the quantities in [13] the coefficient matrix in [3] for the random month model would of the form given below,

let:

$S = [I - M(M'M+\lambda I)^{-1}M']$ ...(adjustment for random herd-year-month)

$Q = [S - SH(H'SH)^{-1}H'S]$ ...(adjustment for fixed effect of herd-year)

then the coefficient matrix will be as follows:

$$\begin{bmatrix} C'QM(A_2-A_1)M'QC & C'QM(A_2-A_1)M'QZ \\ Z'QM(A_2-A_1)M'QC & Z'QM(A_2-A_1)M'QZ \end{bmatrix}$$

where:

M = design matrix for herd-year-months

$\lambda = \sigma_e^2/\sigma_m^2$ on the random month model

The implications of the rolling months models for different herd sizes were also investigated, using herd sizes defined as being either large, moderate or small. The data subsets with the "moderate" herd sizes were simply the data subsets used in the previous parts of this study. To simulate the situations of large herd sizes all herd-year subclasses having less than 8 records were discarded from the initial data subsets. To simulate the situations of small herd sizes each subset was split at random into two, by the animal number. This resulted in a total of 16 subsets.

RESULTS AND DISCUSSION

The average values of the percent increase in the empirical variances on the random month model over the rolling months model showed that the rolling months models with all values of W (i.e. the width for rolling) examined were more efficient than the random month model (Table 7.1). As the value of W increased the rolling months model became slightly more efficient. The plots of the empirical variances for each sire in Figures 7.1-7.7 (for different widths) are consistent with the values of the empirical variances given Table 7.1. The straight lines drawn through the origin in these Figures have a slope of 1.00. The number of sires having empirical variances below the straight line in these figures tended to increase as the value of W increased, showing that the empirical variances on the rolling months model were smaller. All empirical variances plotted in Figures 7.1-7.5 and 7.7 are based on the estimates of sire effects obtained using the estimated values of parameters (i.e. the ratios of variances). However, the empirical variances for the rolling months model plotted in Figure 7.6 are

100

from sire effects estimated using an arbitrary value of λ equal to 50.00. Compared to the estimated value of λ (i.e. 20.60) this value ($\lambda=50.00$) made the rolling months model slightly less efficient. Substantial changes in the value of λ, being equal to either 10 or 35 or 50 or 60 or 70 had only a marginal effect on the empirical variances.

Comparisons of the rolling months models (fitting different widths) with the random month model in terms of the predicted variances of the estimates of sire effects using [10] and [11] are given in Table 7.1. Average increases in the predicted variances of the estimates of sire effects on the random month model, given that the rolling months models with a width either 1 or 2 or 3 or 4 or 5 or 8 months were the true models, were estimated to be respectively 0.77, 1.30, 1.73, 2.01, 2.19 and 2.21%. These values can be interpreted as the percent gain in the efficiency of using the rolling months model over the random month model. Although advantages were small, the rolling months model became more efficient as the width of the rolling months increased. The advantages of using a rolling months model with a width of 8 months over a width of 5 months were not as much as observed for the increases in the widths from 1 to 5 months. Therefore a rolling months model with a width equal to 5 months seems to be optimal. The rolling months model with a value of λ equal to 50 (i.e. an arbitrary value 2.5 times higher than the estimated value of λ with W=5) became closer to the random month model. Its efficiency, in terms of the predicted variances, over the random month model reduced from 2.19% to 0.79%.

The results of the investigation on the usefulness of the rolling months model over the random month model for different herd size situations are given in Table 7.2 and Figures 7.5 and 7.8 - 7.10. The empirical variances on the random month model were larger than those for the rolling months model for all situations. The empirical variances for each sire are plotted for the rolling months model against those for the random month model in Figures 7.5, 7.8 and 7.9 respectively for small, moderate and large herd size situations. More sires have plotted points below the straight line than above, showing that the empirical variances on the rolling months model are smaller in all herd size situations. Considering the predicted variances it was observed that the rolling months model was more efficient for larger herds. The percentage increases in the predicted variances of the estimates of sire effects on the random month model given the rolling months model, from each subset for different herd sizes, are plotted in Figure 7.10. The average values of the observed gains in the efficiency of the rolling months model over the random month model are 1.42, 2.19 and 2.55% respectively for average herd sizes of 4.4, 7.8 and 14.8 records per year. The gains in the efficiency of the rolling months model increase slightly as the herd size increases, however after a herd-year size of about 10 the gains are marginal. Since the correlations between the estimates of sire effects from both models for all herd sizes were very close to unity, however, they would give essentially the same ranking of sires.

The average herd-year size of 14.8 in the data subsets used for large herds in these analyses is incidently equal to the actual number of heifer records available for sire analyses per herd per year in the DPTS herds of the MMB (MMB, 1984). Therefore it can be

concluded that the rolling months model would be about 2.55% more efficient than the random month model (herd-year fixed and herd-year-month random) for sire evaluation in the DPTS. The random month model itself was found to be 37.3% more efficient (in terms of the empirical variances) than the MMB's currently used herd-year-season fixed effects model in Chapter 5.

However, it was concluded that the rolling months procedure did not show as much improvement over the random month model as was imagined it may have. However, the observed and predicted values were consistent. It was observed that there were only 26% of the total number of herd-year-month subclasses filled (in the data subsets for herds of moderate size), and therefore there were not enough connections between months within herd-year to gain the advantages from the variance-covariance structure assumed in the rolling months model. For larger herds there were better connections between months and this was the main reason why the rolling months model was slightly more efficient than the random month model. These results suggest that, compared to the random month model, a model assuming covariances between months would not be advantageous for dairy sires analyses unless herds were large.

SUMMARY

A sire evaluation model assuming covariances among records in the same and different months within herd-year ("rolling months" model) has been compared with the model assuming covariances among records only in the same month and ignoring all covariances between months. The effects of herd-year (fixed, since the data used were on progeny of proven sires), and the overall effects of month of calving

(fixed) and age as a covariable by linear and quadratic regressions were the same in both models. The data comprised 8 independent subsets of the first lactation fat yield records on 49242 progeny of 69 widely used proven Holstein-Friesian sires in 1628 herds in England and Wales. The rolling months model, assuming linear covariance between months up to 5 months apart and the rest of the covariances between months zero, was observed to be only 2.19% more efficient than the random month model. These data had an average herd-year size of 7.8. Analyses for different herd size situations showed that this model was 1.42 and 2.55% more efficient for herds with 4.4 and 14.8 records per year. However it was concluded that this model would not be advantageous over the random month model for dairy sires analyses as the average herd-year size in most situations is less than 14.8.

TABLE 7.1. Percent increases in the empirical, and predicted variances of the estimates of sire effects on the random month model given that the rolling months models with different widths (W) are the true models.

Subset	W=1 [†]$\lambda=4.12$	W=2 $\lambda=8.24$	W= 3 $\lambda=12.36$	W= 4 $\lambda=16.48$	W= 5 $\lambda=20.60$	W= 5 [ψ]$\lambda=50.00$	W=8 $\lambda=32.95$
1	0.75	1.25	1.63	1.92	2.05	0.74	2.11
2	0.86	1.48	1.93	2.25	2.42	0.87	2.32
3	0.82	1.36	1.89	2.23	2.46	0.89	2.54
4	0.76	1.27	1.60	1.85	1.99	0.76	2.05
5	0.69	1.07	1.41	1.60	1.77	0.68	1.76
6	0.76	1.27	1.65	1.87	2.05	0.71	1.95
7	0.74	1.26	1.79	2.12	2.33	0.75	2.36
8	0.78	1.42	1.92	2.23	2.46	0.89	2.58
Average increase	0.77	1.30	1.73	2.01	2.19	0.79	2.21
[¶]Emp.var.	2.29	3.19	3.47	3.78	3.95	3.40	4.29

[†]$\lambda = \sigma_e^2/\phi$ added to the diagonal of the P'P submatrix in rolling months model. The value of λ as the ratio of residual variance to variance of records in the same month (σ_e^2/σ_m^2) used in the random month model was 8.07. The σ_e^2 and the variance of records in the same month (C_0) respectively were 690 and 168 kg^2.

[ψ]The value of λ equal to 50.00 is an arbitrary value.

[¶]Average values of the % increase in empirical variances for random month model over those for the rolling months model.

TABLE 7.2. Percent increases in the empirical and predicted variances of the estimates of sire effects on the random month model given that the rolling months model with $\bar{W}=5$ and $\lambda = 20.60$ is the true model, for all situations with different herd sizes.

[†]Herd size	No. of subsets	Average herd-year size	% increase in the empirical variances on the random month model over the rolling months model	% increase in predicted variance on random month model	[¶]Product moment correlation
Small	16	4.4	1.32	1.42	0.997
Moderate	8	7.8	3.95	2.19	0.994
Large	8	14.8	1.33	2.55	0.994

[†]The data on widely used proven sires are regarded as being from herds of moderate size. Small herds were created by splitting each subset at random into two, and large herds by discarding all herd-year subclasses having less than 8 records.

[¶]Product moment correlation between the estimates of sire effects on the rolling months and the random month models.

Fig. 7.0. The trends in the various months apart
observed covariances and the alternative
trends investigated; ——— observed,
− − − investigated.
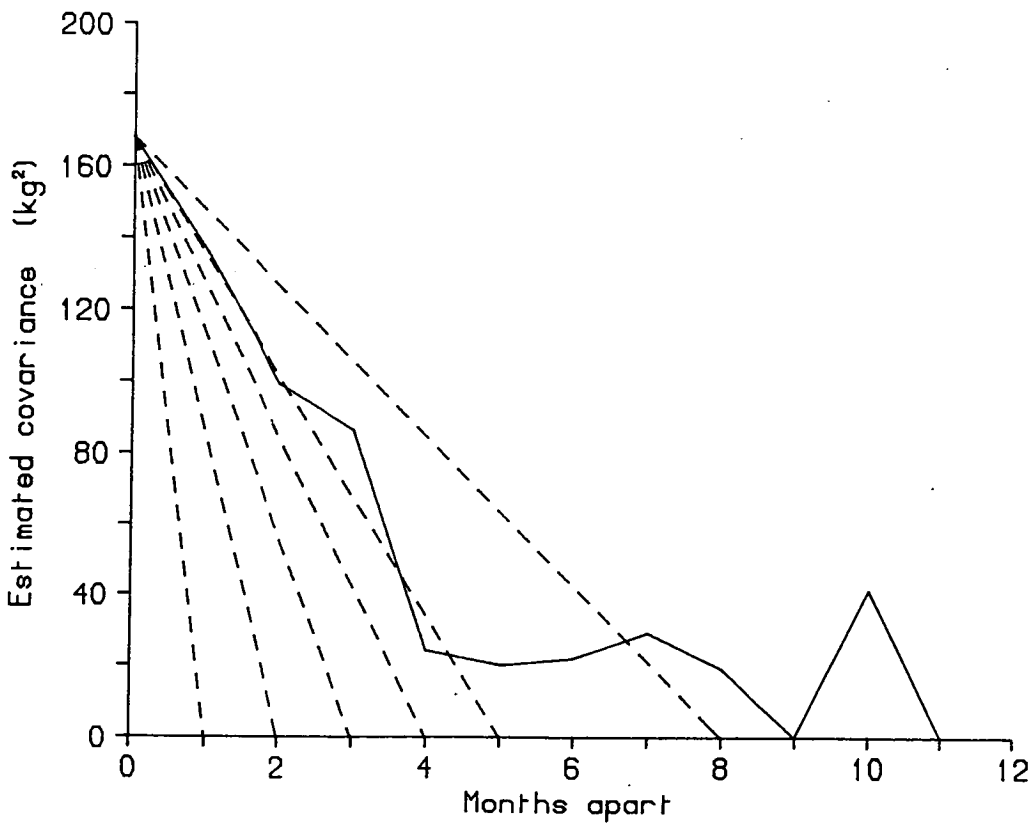
Fig 7.1. Empirical variances for each sire plotted for the rolling months model (width [W] = 1 month) against those for the random month model. The straight line has a slope of 1.00 (k = ratio of residual variance to the variance of records in the same pseudo-month)
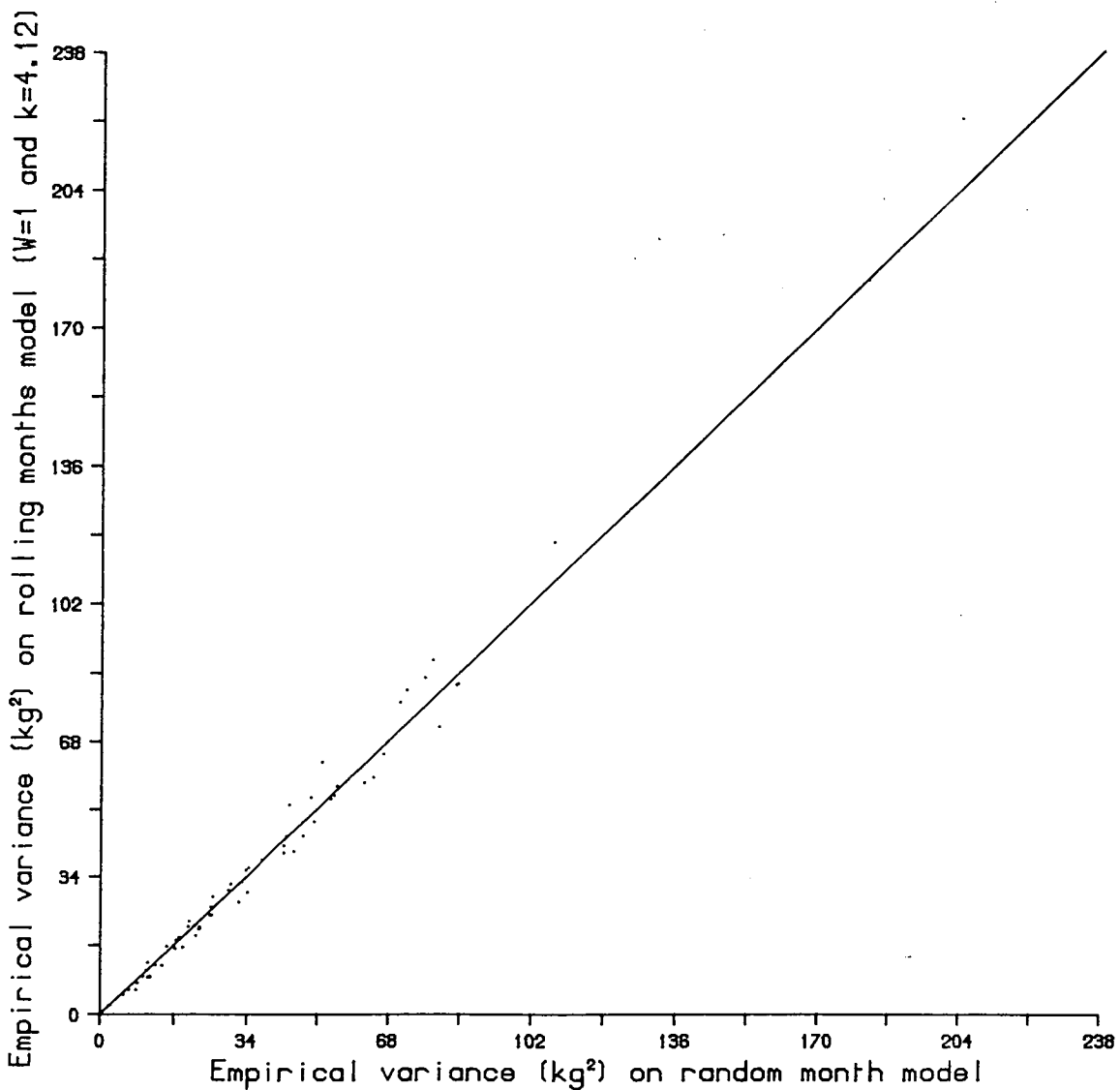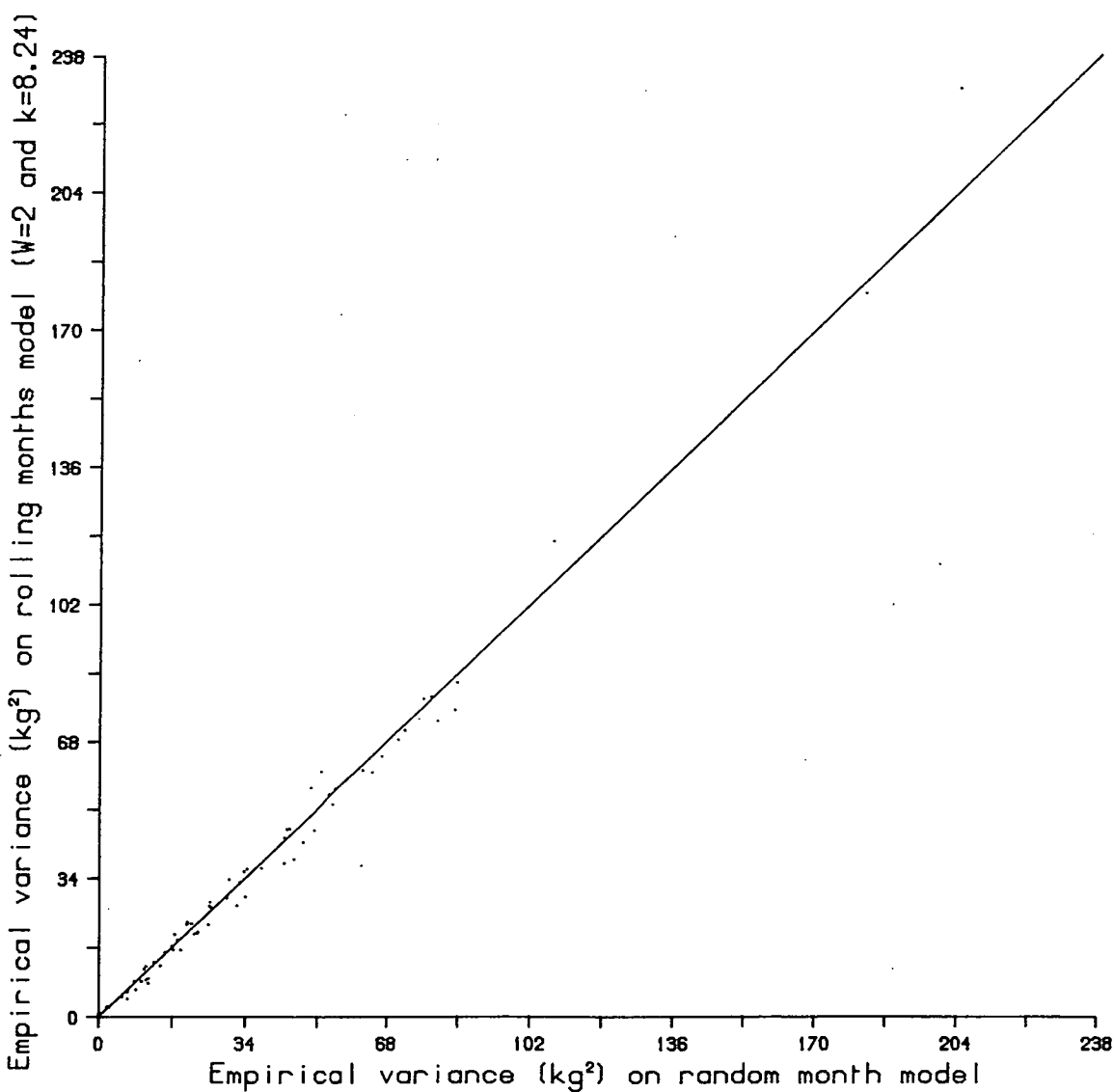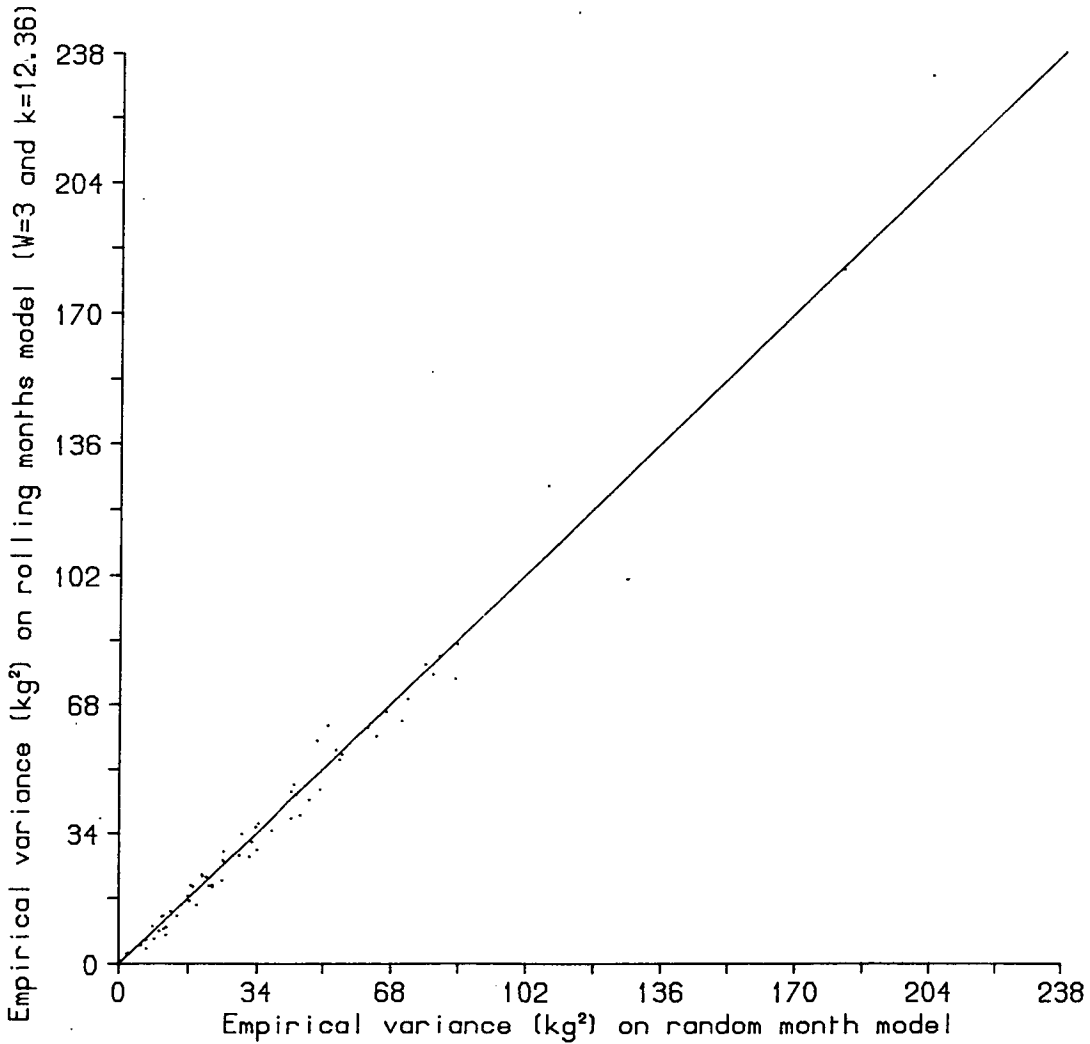
Fig 7.2. Empirical variances for each sire plotted for the rolling months model (width [W] = 2 months) against those for the random month model. The straight line has a slope of 1.00 (k = ratio of residual variance to the variance of records in the same pseudo-month).
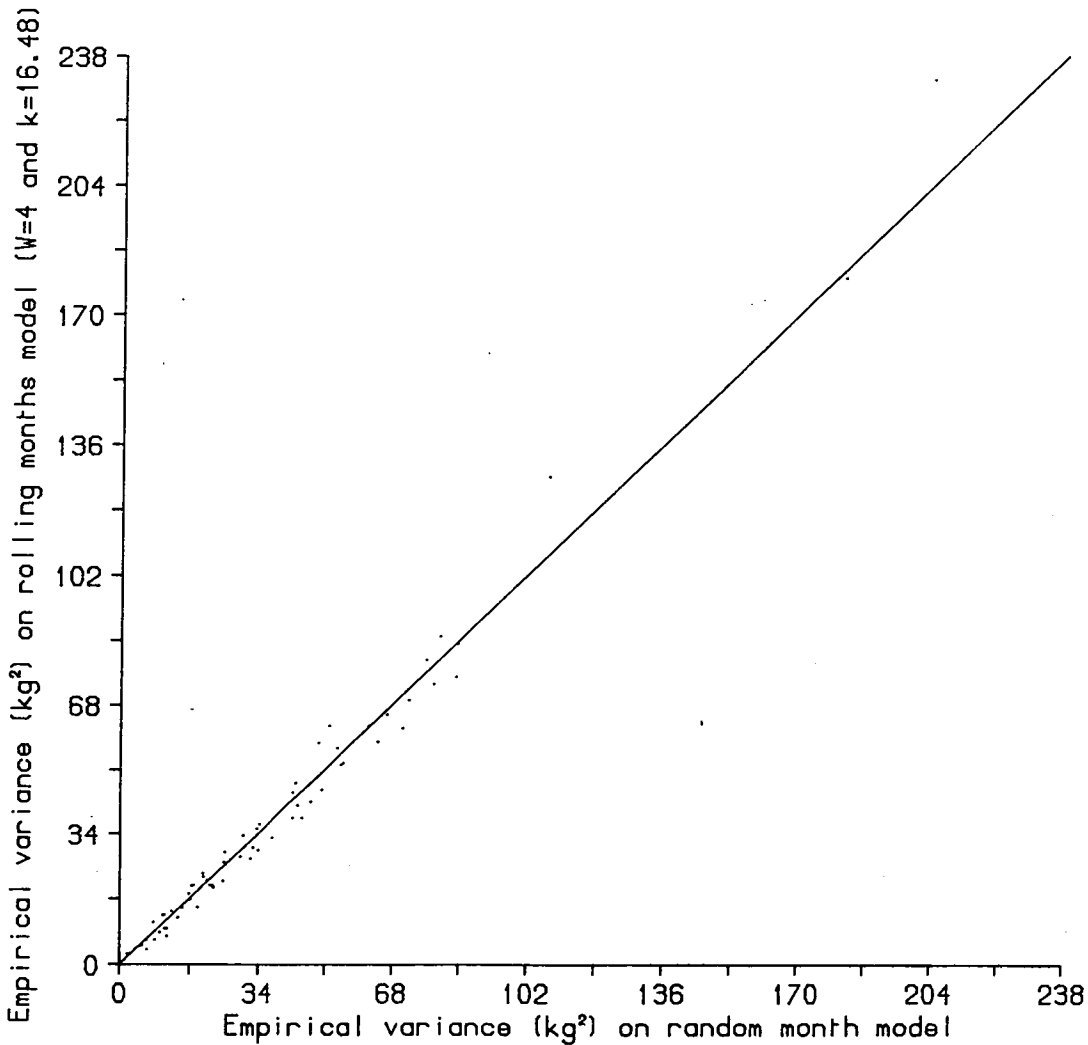
Fig 7.3. Empirical variances for each sire plotted for the rolling months model (width [W] = 3 months) against those for the random month model. The straight line has a slope of 1.00 (k = ratio of residual variance to the variance of records in the same pseudo-month).

Fig 7.4. Empirical variances for each sire plotted for the
rolling months model (width [W] = 4 months) against
those for the random month model. The straight line
has a slope of 1.00   (k = ratio of residual variance
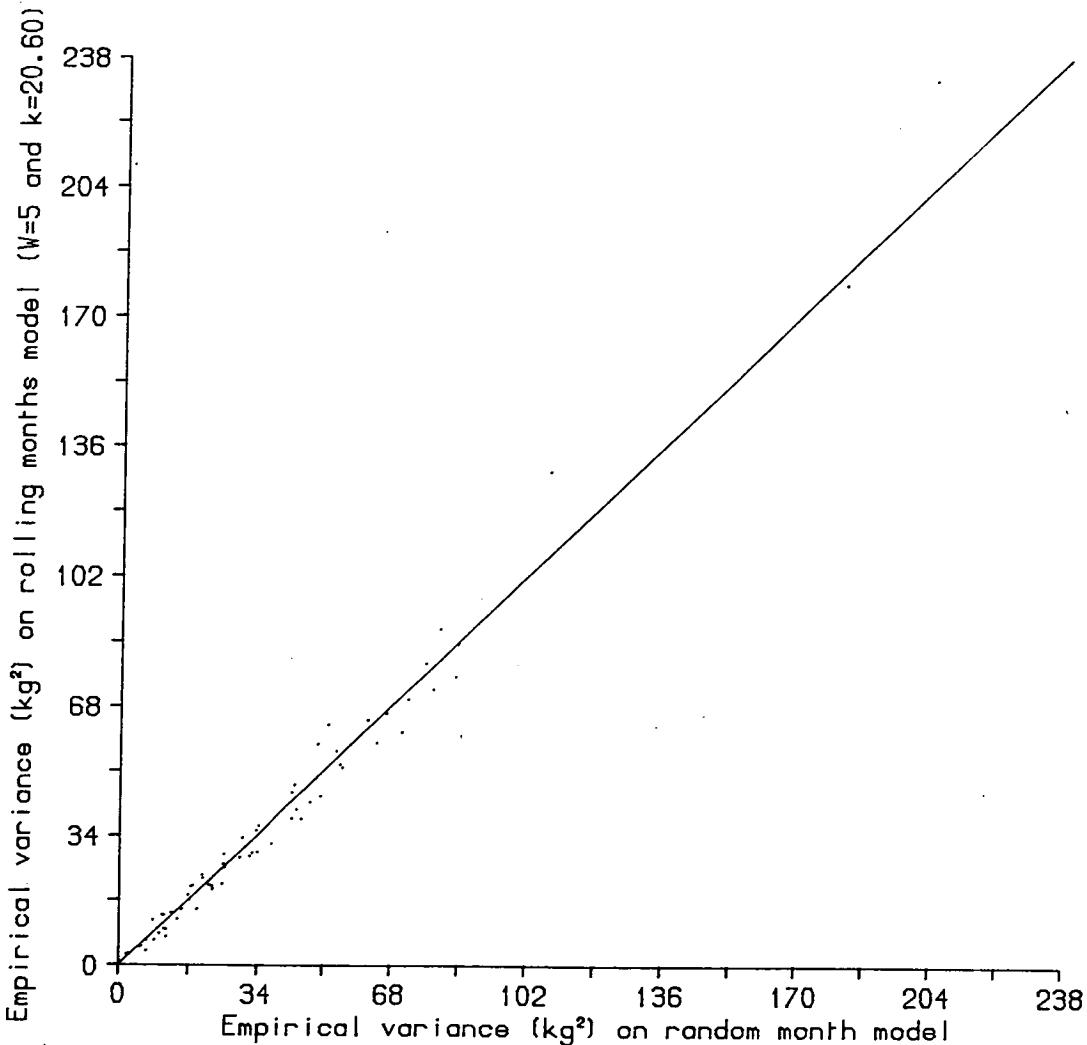to the variance of records in the same pseudo-month).

Empirical variance (kg²) on rolling months model (W=4 and k=16.48)

Empirical variance (kg²) on random month model

Fig 7.5. Empirical variances for each sire plotted for the rolling months model (width [W] = 5 months) against those for the random month model. The straight line has a slope of 1.00 (k = ratio of residual variance to the variance of records in the same pseudo-month).
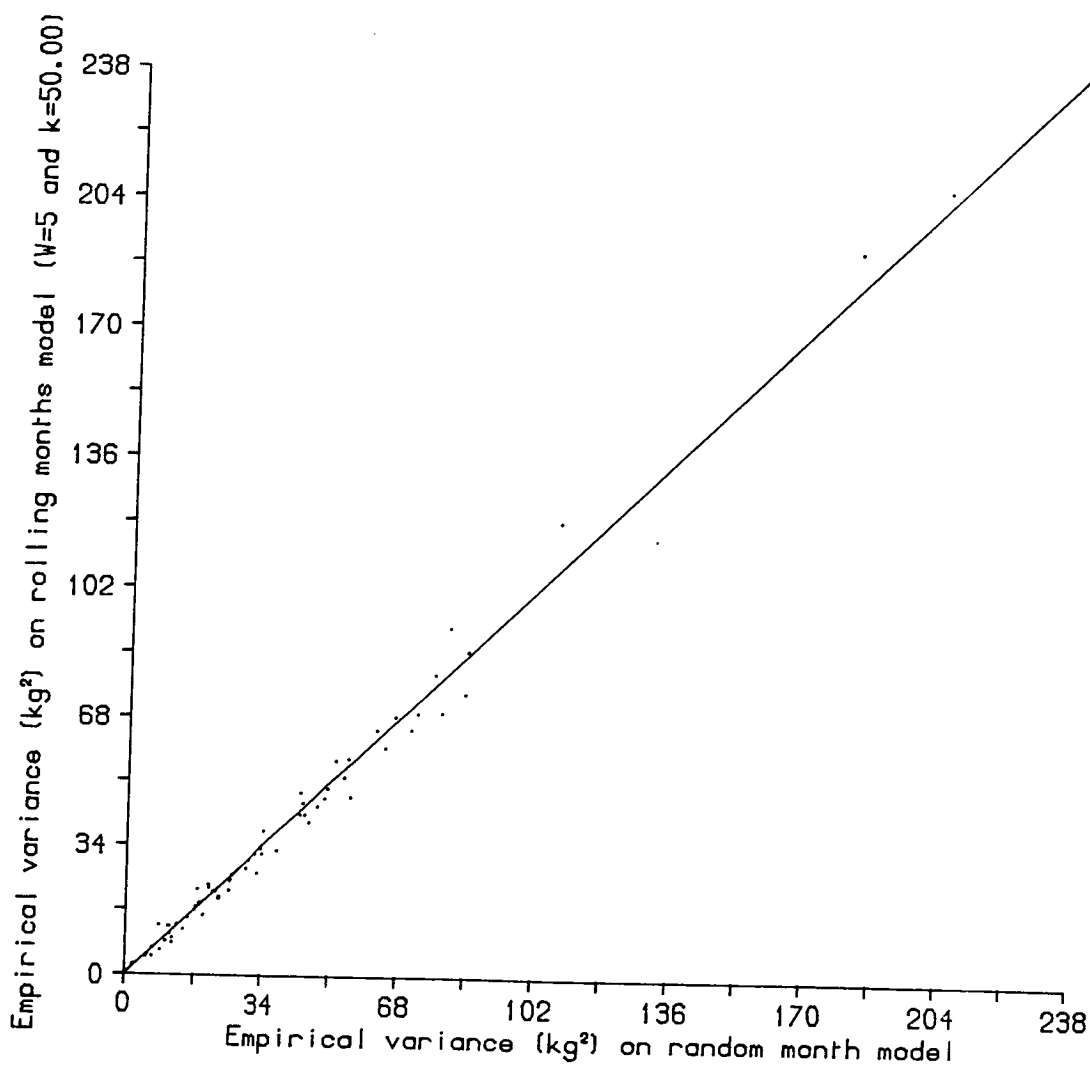
Fig 7.6. Empirical variances for each sire plotted for the rolling months model (width [W] = 5 months) against those for the random month model. The straight line has a slope of 1.00 (k = ratio of residual variance to the variance of records in the same pseudo-month).

Fig 7.7. Empirical variances for each sire plotted for the rolling months model (width [W] = 8 months) against those for the random month model. The straight line has a slope of 1.00 (k = ratio of residual variance to the variance of records in the same pseudo-month).
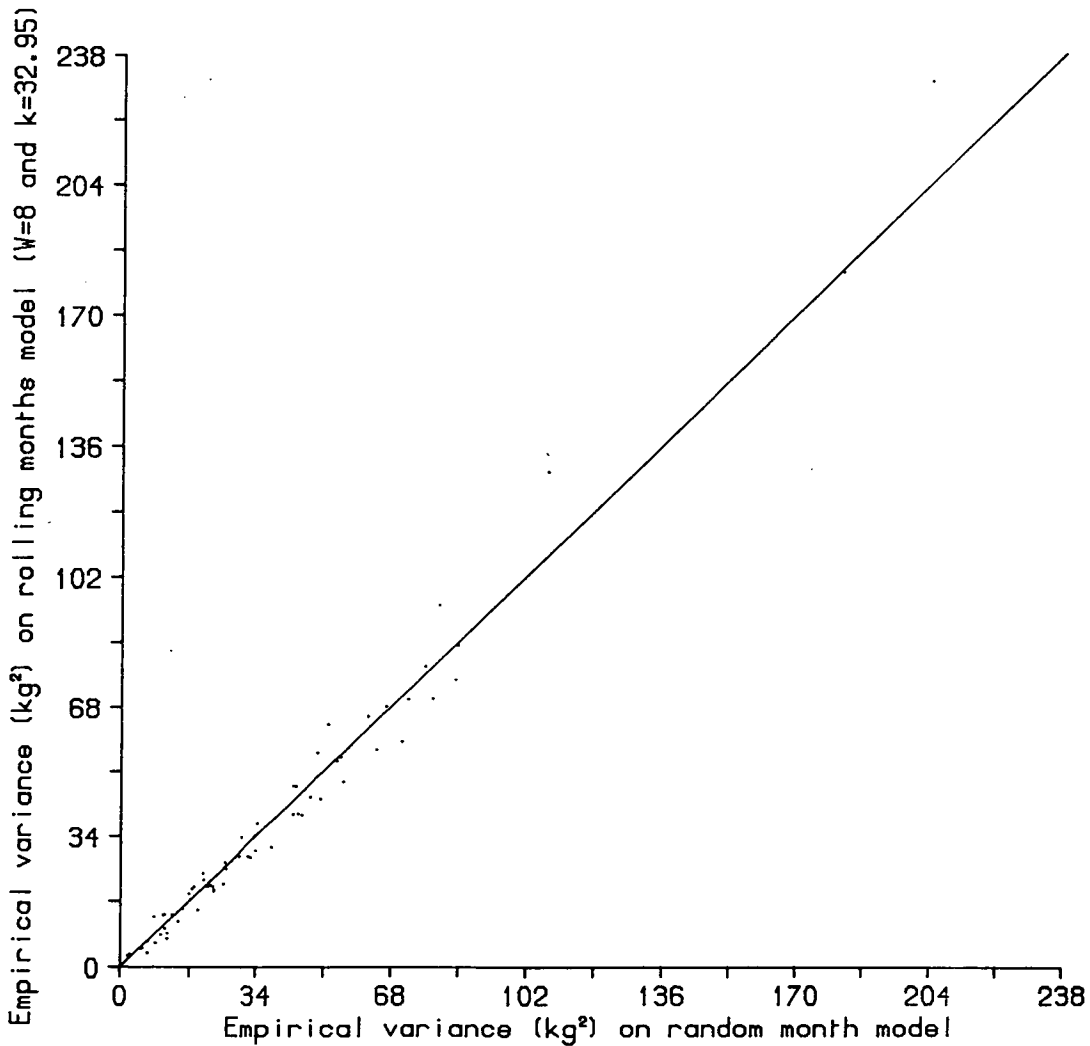
Fig 7.8. Empirical variances for each sire plotted for the rolling months model (width [W] = 5 months) against those for the random month model, for small herds. The straight line has a slope of 1.00 (k = ratio of the residual variance to the variance of records in the same pseudo-month).
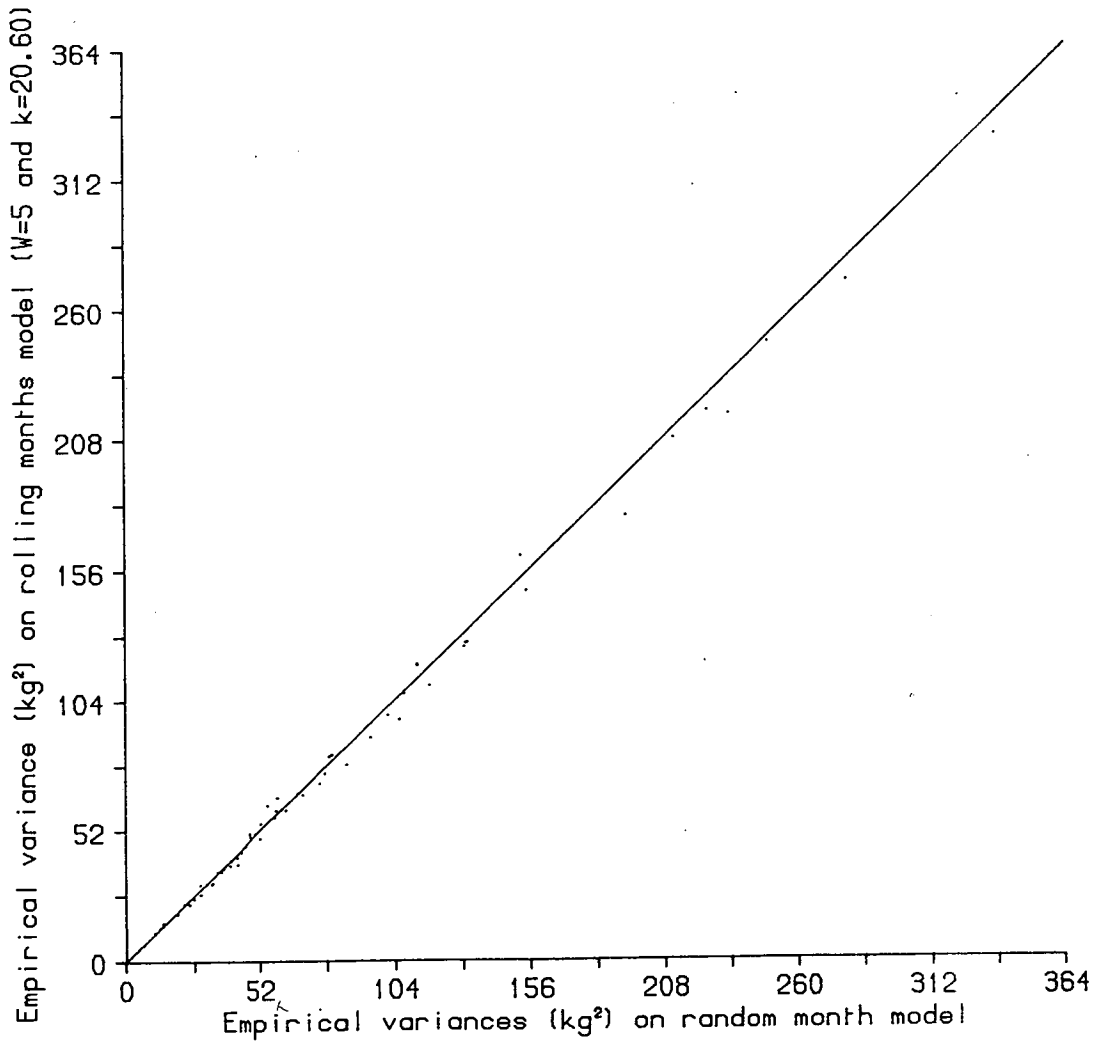
Fig 7.9. Empirical variances for each sire plotted for the
rolling months model (width [W] = 5 months) against
those for the random month model, for large herds.
The straight line has a slope of 1.00 (k = ratio of
the residual variance to the variance of records
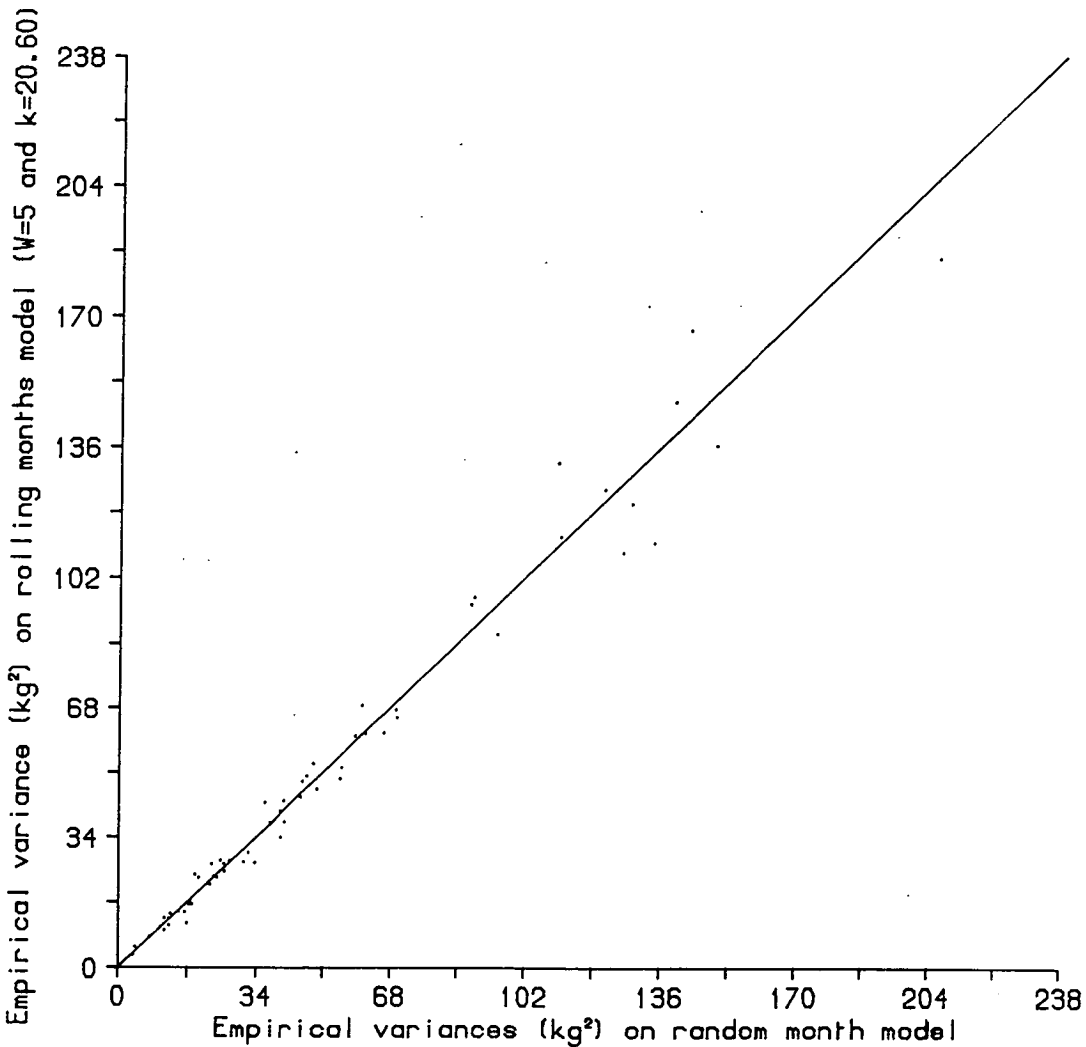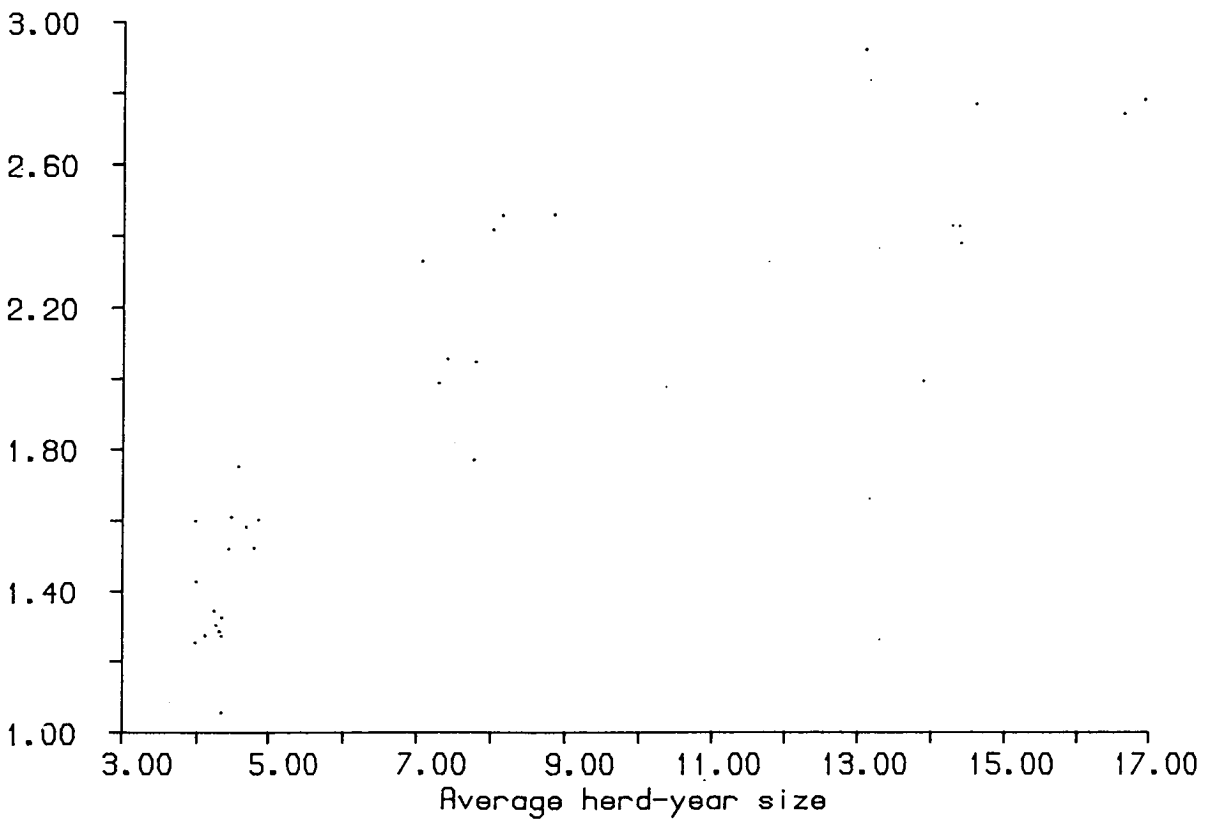in the same pseudo-month).

Fig 7.10. Percent increase in the predicted variance
of sire effects for different herd-year
sizes on the random month model given that
rolling months model (width = 5 months)
is the true model. The ratio of residual
variance to the variance of records in the
same pseudo-month used for these analyses
was 20.60.

Chapter 8

GENERAL DISCUSSION AND IMPLICATIONS

DISCUSSION

Of the several questions asked in these studies the main ones were on the season effects as follows:

(1) how should the months of calving be grouped to form seasons?

(2a) what is the effect of considering season effects to be random rather than fixed?

(2b) how does the length of season affect the accuracy of sire evaluation using models with season effects fitted as random, as opposed to models with season effects as fixed?

(3) are there any covariances between months, and would a model accounting for such covariances, if there are any, be helpful in sire evaluation?

There do not seem to be any specific guide lines for grouping months into seasons. However, the smallest herd-year X month within season interaction was considered to be a useful criterion in sire analyses for using milk production records precorrected for month effects. Other criteria used were the residual variance and the effective number of daughters (i.e. the number of progeny adjusted

for herd-year-season effects). Several analyses fitting seasons of variable lengths, having different trends in production, showed little differences in herd-year X month within season interactions and residual variances. Comparing different seasonal groupings with three seasons, the grouping currently used by the MMB (December-March, April-July and August-November) was found to be optimal. A two seasons grouping (February-July and August-November) increased the number of within group comparisons, but it gave a larger herd-year X month within season interaction.

Using a model fitting seasons of either 1, 2 or 4 months it was observed that the correlations among records tended to be larger when seasons were shorter. However, although shorter seasons were therefore considered to be useful for progeny group comparisons, the losses of information due to smaller subclasses were substantial when herd-year-season effects were fitted as fixed. From the comparisons of herd-year-season fixed effects models with seasons of different lengths (see Table 5.2 in Chapter 5) it was observed that the estimated variances of sire effects were larger by 14% with a season of 2 months, and by 36% with a season of one month, than those for a model with each season of 4 months (currently used by the MMB). The corresponding average values for the empirical variances using seasons of either 1 or 2 or 4 months were respectively 79, 72 and 57 kg^2, showing that the herd-year-season fixed effects models with shorter seasons were considerably less efficient. Models fitting the effects of year within herd and season within herd-year, or just season within herd-year, as random effects, were observed to be more efficient than the herd-year-season fixed effects model. The use of random effects models increased the accuracy of sire evaluation, not just from the

recovery of the inter-season information but also from the use of groupings of shorter duration i.e. one month. Compared to the MMB's currently used herd-year-season fixed effects model (each season of 4 months) the following two models: (i) herd-year fixed and herd-year-month random, and (ii) herd fixed, herd-year and herd-year-month random, gave respectively 16 and 32% smaller variances of sire effects. The respective average values for the empirical variances were 44 and 37 kg^2, compared to 57 kg^2 for the herd-year-season fixed effects model. Much of the loss of information involved in using shorter seasons can be avoided by fitting them as random rather than fixed effects. As the seasons were made shorter the advantages of using random effects models became larger, suggesting that seasonal grouping could be reduced to a single month (N.B. the information necessary to subdivide even further than a month were not available in the data set). Consequently, the models fitting season of one month as a random effect would be helpful as they would account for the herd-year-season X month interaction effect. Therefore the random effects models eventually overcome the limitations of seasonal groupings discussed in Chapter 3.

In contrast to the assumptions of zero covariances between months within season or across seasons within herd-year in the herd-year-season fixed effects model, the covariances between months up to 5 months apart within herd-year were found to be considerably larger than zero, with an almost linear decline from 0 through to 5 months apart. All covariances beyond 5 months apart were small. In order to account for these observed covariances a model – the "rolling month" model – was fitted in which herd-year effects were regarded as fixed and the covariances from 0 to 5 months apart were

assumed to decline linearly as the distance between months increased. All covariances beyond 5 months apart were assumed to be zero. This model did not show as much improvement over the random month model (herd-year fixed and herd-year-month random) as was imagined it may have. The random month model takes account of only the covariance among records in the same month (i.e. the 0-month-apart covariance) and all between months covariances are assumed zero. It was concluded that there were not enough herd-year-month subclasses filled to provide the connections between months necessary to gain the advantages from the variance-covariance structure assumed in the rolling months model. However, this procedure was found to be slightly more efficient than the random month model for larger herds where the connections between months were better.

In addition, a "herd-year-month-age" model was also examined in view of the findings of Barker and Robertson (1966) that the residual mean squares within herd-year-month-age subclasses were considerably smaller than those within herd-year-month subclasses. Least squares analyses were conducted on two data subsets (25158 and 24084 records, as described in Chapter 3) using a model fitting the effects of herd-year, month within herd-year, age class within herd-year-month, sire and age as a covariable by linear and quadratic regressions. Based on the age-milk yield relationship in British Holstein-Friesian shown by McClintock (1982) three age classes (less than 25 months, 25-30 and over 30 months) were fitted. The variance components for herd-year, month within herd-year, age class within herd-year-month, and residual variances for milk and fat yields, using the procedure shown in Table 4.2, are given in Table 8.1. Since the estimates of variance components for age

within herd-year-month are close to zero, herd-year-month X age interactions can be ignored. The residual variances for milk and fat yields on herd-year-month plus age model were not different from those on a herd-year-month-age plus age model. The results from both subsets were consistent on untransformed as well as a log scale (see Table 8.1). The residual mean squares within herd-year-month subclass reported by Barker and Robertson (1966) were higher, presumably because of overall age effects. Assuming the same covariance among records in a herd-year-month or a herd-year-month-age subclass, the empirical variances on two models: (1) herd-year fixed, herd-year-month random and age as a covariable, and (2) herd-year fixed, herd-year-month-age random and age as a covariable, are plotted in Figure 8.1. The average values of empirical variances for models 1 and 2 were quite similar, being 44.3 and 44.5 kg^2, respectively. Considering the additional interactions accounted for in the latter model, it was concluded that the interactions of age with either herd, year, or month of calving do not seem to bias the sire evaluations, in contrast to the reports of Miller et al (1970). The product moment correlations between the sire effects estimated using these two models were not different from unity.

In all models investigated in these studies the effect of sire was fitted as a fixed effect because the data used were on widely used proven sires. However, if the data were on young sires and sire effects were fitted as random, the inferences about the comparison of models would presumably have been the same. The data on widely used sires were useful in making the empirical comparisons.

Sire evaluation models for very small herds, with particular reference to India:

In contrast to the dairy cattle population structure in Britain the herds in India are very small. Therefore some more investigations were carried out to determine a model suitable in small herd situations.

The same 8 subsets of data as described in Chapter 5 were used to simulate small herds situations. Each of these subsets was further split at random by animal number into two smaller subsets in Chapter 7, resulting in a total of 16 subsets. The average herd-year size in these subsets was 4.35 and the herd-year-month size 1.74 records (Table 8.2). The data on 63 sires (48101 records) having records in all 16 subsets were used.

Based on the findings in Chapter 5 on the usefulness of different models of sire evaluation, it would not be appropriate to use a model with herd-year-season effects as fixed because this model results in heavy losses of information when subclass sizes are small. A model assuming a variance-covariance structure between months (Chapter 7) would also not be helpful or advantageous, since the gains of using this model over a model fitting herd-year fixed and herd-year-month random were observed to be very small for small herds. In view of these considerations the following two models were examined using the above data from British herds used to simulate a small herds situation:

Model 1 : $Y_{ijklm} = HY_{ij} + M_{ijk} + Z_l + bA + b'A^2 + e_{ijklm}$

Model 2 : $Y_{ijlm} = HY_{ij} + Z_l + bA + b'A^2 + e_{ijlm}$

where:

Y_{ijklm} = record of the mth progeny of the lth sire calved at age A

in the jth year and the kth month of the ith herd,

Y_{ijlm} = record of the mth progeny of the lth sire calved at age A

in the jth year of the ith herd,

HY_{ij} = joint fixed effect of the ith herd and the jth year

M_{ijk} = random effect of the kth month of calving in the jth year

of the ith herd,

Z_l = fixed effect of the lth sire (fixed effect since proven sires),

b and b' = linear and quadratic regression coefficients of fat yield

on age at calving A.

e_{ijklm} = random error in model 1,

e_{ijlm} = random error in model 2.

The ratio of residual variance to the variance of records in a herd-year-month suclass (σ_e^2/σ_m^2) was 8.07 as given in Chapters 4 and 5. The data were precorrected for the overall effects of month of calving. Alternatively the effect of month of calving could be fitted in the model as a fixed effect. Similarly, age effects could have been precorrected. These two models were compared using the empirical variances of sire effects as in Chapters 5 and 7. This criterion is useful because an efficient model should give similar estimates of sire effects from independent subsets of data.

The residual variances after fitting the above two models to different data subsets are given in Table 8.2. These variances were consistent for both models, showing that the data split was reasonable. The empirical variances for model 1 (herd-year fixed, and herd-year-month random) and model 2 (herd-year fixed ignoring herd-year-month) were respectively 98.7 and 99.9 kg^2. Although the averages of the empirical variances using both models are similar, more sires have a smaller empirical variance on model 1 than on

model 2 (Figure 8.2), suggesting that model 1 is the more appropriate.

It could be argued, however, that the herd-year and herd-year-month sizes in India are not even as big as in the data created for small herds from British herds. Most farmers own a single milking cow, with the average herd size for sire analyses being under two. The main breeding practices comprise the crossbreeding of indigenous cattle with Holstein-Friesian, Jersey and Brown Swiss. A large quantity of semen of proven bulls of exotic breeds is imported to produce bulls of exotic breeds and crossbred bulls at the institutional herds. The main criterion for choosing young bulls for use in field is the dam's yield and their sire's merit. However, in a few breeding schemes progeny testing has also been recently introduced. These schemes require a model of sire evaluation that can make the best use of the information. In view of the above breeding practices the evaluation of the crossbred as well as purebred exotic young bulls may be useful. Apart from the common effects of herd, year, season, age and any interactions among them, the effects of exotic inheritance level of sire and of the dam of progeny are the known important sources of variation in milk production. Therefore, the grouping of bulls according to exotic inheritance levels (Hagger and Dempfle, 1983) along with other criteria, e.g. year of birth, stud, etc. may be helpful for effective comparison of bulls. At present, the artificial insemination is mostly undertaken by the village milk producers' cooperatives. It is true that the bigger the village cooperative the better the facilities for breeding. Therefore, similar to the larger herd situations it may be useful to assume the village effects to be fixed, in order to remove any possible bias in

evaluations due to non-random use of bulls across villages. Since the yearly trends in milk production within villages would seem to be large due to increasing facilities for dairy cattle improvement, it may be more appropriate to assume the village-year effects to be fixed rather than just the village effects fixed.

The structure and design of Indian field data were not available to simulate data true to the Indian situation. Ignoring the effects of exotic inheritance levels or the breed of bulls and the dam of progeny, and considering the above results for sire analyses in small herds, the possibilities of a model suitable for India are discussed below. Assume the number of cows in a herd-year subclass in British herds to be analogous to a group of cows recorded in a village in India in one year, and likewise a herd-year-month subclass to be analogous to a farmer nested within village-year. If the farmer effects are considered analogous to the herd-year-month effects it has to be assumed that all cows of a farmer calving in the same or different months have equal covariance with each other. To overcome this problem, farmer-month effects could have been assumed to be analogous to the herd-year-month effects, however this would have resulted in most subclasses being of size one. The error variance-covariance between farmer effects within village-year can not be assumed to be analogous to those of herd-year-month effects within herd-year, however these results of sire analyses for small herds do give an indication of the consequences of ignoring an effect. The variation between farmers within village would presumably be much larger than the variation between months within herd-year in British data. Therefore, a model ignoring the farmer effects may result in considerably larger variances of the estimates of sire effects, and consequently a bias in the ranking of sires.

126

In view of these results and the dairy cattle population structure in India, a model with the following effects may be suitable: fixed effect of village-year, random effect of farmer within village-year, fixed effect of group of sire, random effect of sire within group, fixed effects of month of calving and age at calving, and fixed effect of breed of the dam of progeny. The effects of age and month of calving could be precorrected if appropriate correction factors are derived from large data sets. The true model, however, should be based on adequate investigations using actual data sets.

IMPLICATIONS

Considering that yearly trends in milk production within herd are an important source of bias in sire evaluations in most breeding schemes, a model in which herd-year effects are regarded as fixed and herd-year-season random (season being one month duration in this model) would be more efficient than a herd-year-season fixed effects model, as the random (season) effects model will recover information from inter-season comparisons. If it is necessary to fit herd-year-season effects as fixed (e.g. for the existance of seasonal trends within herds, convention, etc) then substantial gains in the accuracy of sire evaluation could most probably still be obtained by fitting herd-year-month as a random effect along with herd-year-season (fixed). If the connections between months within herd are sufficient then a model assuming covariance between months will be useful.

In view of the findings in these studies based on the data from the herds participating in the Dairy Progeny Testing Scheme of the MMB, it can be suggested that a model with herd-year effects as

127

fixed and herd-year-month effects as random will be considerably more efficient than their currently used herd-year-season fixed effects model.

Although it does not seem necessary to fit herd-year-seasons as fixed in the DPTS, investigations on a model with herd-year-season effects as fixed and herd-year-month random may have been useful. Since the random effects model showed a larger advantage with shorter seasons, it may have been worth examining a model with a season shorter than one month. Similarly, some models could have been examined in which herd-period effects (period of 2 years) are fixed, and herd-period-season (season of 4 months) and herd-period-month effects are random.

Most importantly, the random effects models were observed to be of great value for small herds. Considering the environmental effects strictly in terms of herd, year, season etc., a model with random effects in a nested structure would be very helpful for small herd size situations, e.g. India.

TABLE 8.1. Herd-year-month-age analyses for milk and fat yields (kg).

[†]Source	DF	Milk yield [¶]F-ratio untrans-formed	Milk yield [¶]F-ratio log scale	Milk yield [ψ]variance compo-nent(%)	Fat yield [¶]F-ratio untrans-formed	Fat yield [¶]F-ratio log scale	Fat yield [ψ]variance compo-nents(%)
Data set (25158 records):							
Herd-year	3232	5.329	5.304	38.1	6.620	6.644	45.0
Herd-year-month	7110	1.191	1.156	5.0	1.234	1.209	5.4
Herd-year-month-age	2212	1.019	1.083	0.6	1.036	1.078	1.0
Residual	12533			56.3			48.6
Residual variance (kg^2)			480902			686	
Residual variance ignoring herd-year-month-age			482243			690	
Data set (24084 records):							
Herd-year	3112	5.543	5.654	38.4	6.542	6.658	44.7
Herd-year-month	6607	1.259	1.220	5.7	1.284	1.237	5.5
Herd-year-month-age	2133	0.977	1.011	0.0	0.969	1.003	0.0
Residual	12161			55.9			49.8
Residual variance (kg^2)			475095			695	
Residual variance ignoring herd-year-month-age			473445			692	

[†]In all analyses sire effect fitted as a fixed effect and age as a covariable by linear and quadratic regressions, and the overall effect of month of calving removed by precorrection of data.

[¶]herd-year tested against herd-year-month, herd-year-month against herd-year-month-age, and herd-year-month-age against residual variance.

[ψ]given as a percentage of the total variance due to herd-year, herd-year-month, herd-year-month-age and residual variance, negative components were set to zero.

TABLE 8.2. Average herd-year and herd-year-month sizes, and residual
mean squares (R.M.S.) for fat yields from different subsets
simulated for small herds.

Subset	Total no. records	Herd-year size	Herd-year-month size	[†]R.M.S. (kg^2)	
				model 1	model 2
1	2822	4.26	1.69	681	742
2	3269	4.43	1.80	717	783
3	3266	4.69	1.83	690	754
4	2475	4.11	1.72	659	717
5	2806	4.33	1.68	675	735
6	3089	4.23	1.75	703	766
7	3180	3.99	1.64	706	770
8	3245	4.81	1.85	708	770
9	2777	4.33	1.70	694	753
10	3237	4.46	1.78	681	739
11	3277	4.56	1.77	735	802
12	2414	3.98	1.70	679	742
13	2771	4.33	1.68	664	719
14	3122	4.30	1.76	676	737
15	3095	3.98	1.67	707	774
16	3256	4.87	1.85	681	741
Average	3006	4.35	1.74		

[†]In model 1 herd-year is fixed and herd-year-month random
whereas in model 2 herd-year effects are fitted as fixed
and herd-year-month effects are ignored.

Fig 8.1. Empirical variances for each sire plotted
for a model with herd-year (HY) fixed and
herd-year-month-age (HYMA) random effects
against those for a model with herd-year
fixed and herd-year-month (HYM) random
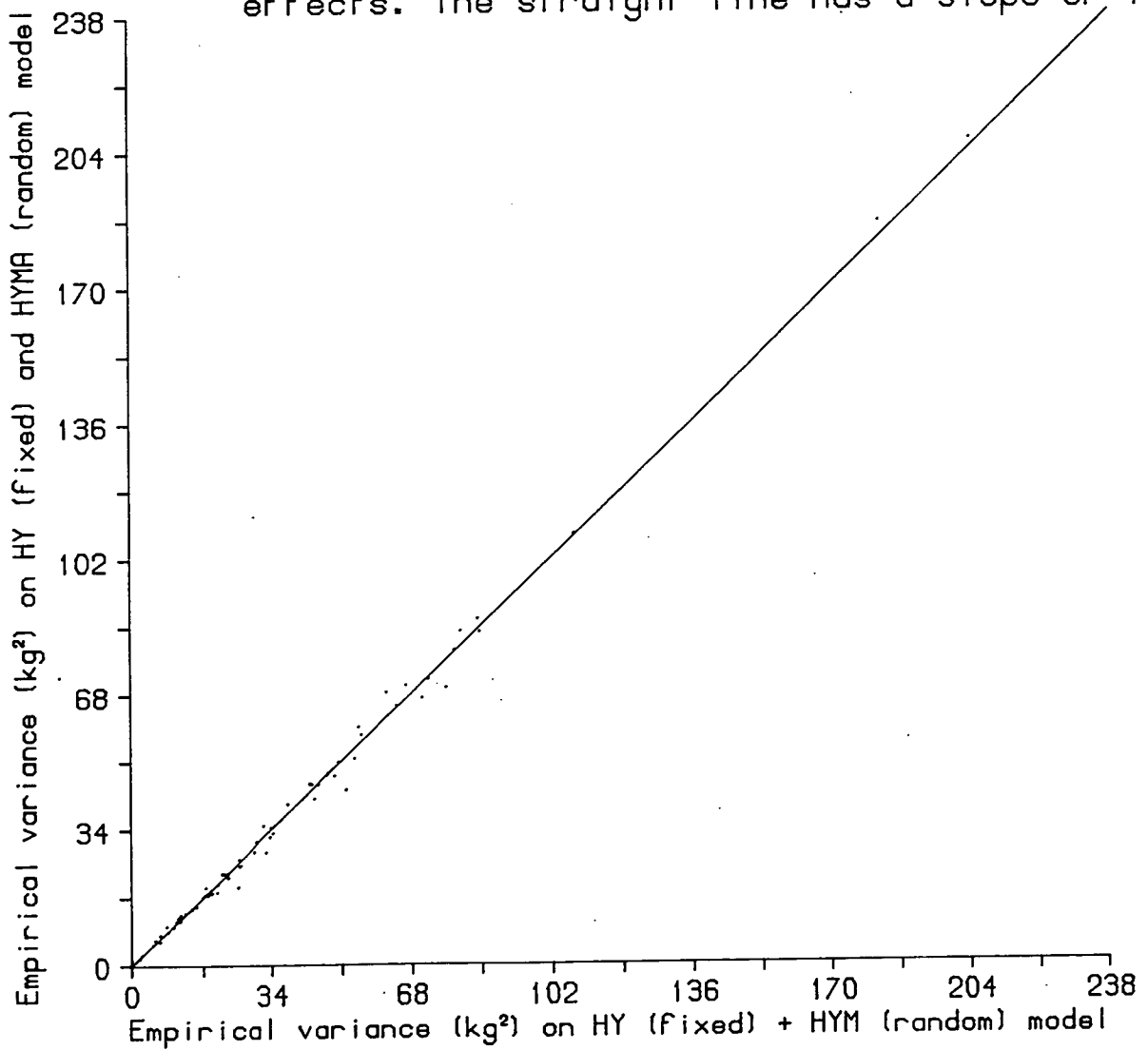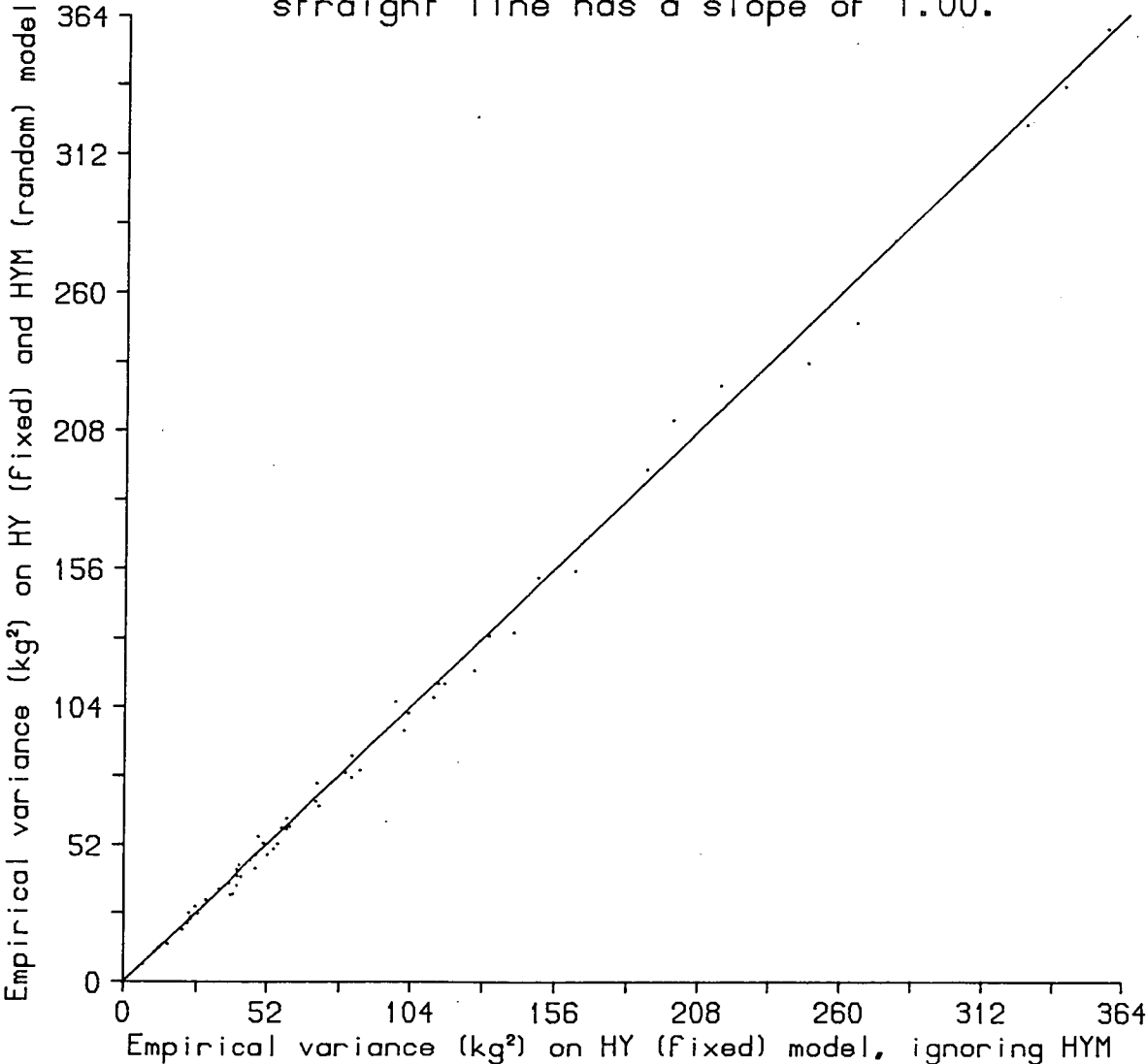effects. The straight line has a slope of 1.00.

Fig. 8.2. Empirical variances for each sire plotted for a model with herd-year (HY) fixed and herd-year-month (HYM) random effects against those for a model with herd-year fixed, ignoring HYM. The straight line has a slope of 1.00.

BIBLIOGRAPHY

Allaire, F.R. and Gaunt, S.N. 1965. First lactation contemporary

comparisons as indicators of environmental influences on

on daughter records used for sire evaluation.

J. Dairy Sci. 48:454-461.

Arora, K.K. and Freeman, A.E. 1971. Environmental correlations between

paternal half-sisters for milk and milk fat production.

J. Dairy Sci. 54:880-885.

Auran, T. 1973. Studies on monthly and cumulative monthly milk yield records.

1. Effect of age, month of calving, herd and length of first test period.

Acta Agric. Scan. 23:189-199.

Barker, J.S.F. and Robertson, Alan. 1966. Genotypic and phenotypic

parameters for the first three lactations in Friesian cows.

Anim. Prod. 8:221-240.

Bartlett, M.S. 1978. Nearest neighbour models in the analysis of

field experiments. J. Roy. Stat. Soc. B 40:147-174.

Bereskin, B. and Freeman, A. E. 1965. Genetic and environmental factors in

dairy sire evaluation. I. Effects of herds, months and year-season on

variance among lactation records; repeatability and heritability.

J. Dairy Sci. 48:347-351.

Bereskin, B. and Lush, J.L. 1965. Genetic and environmental factors

in dairy sire evaluation. III. Influence of environmental and

other extraneous correlations among the daughters.

J. Dairy Sci. 48:356-360.

Bolgiano, D.C., Van Vleck, L.D. and Everett, R.W. 1979. Fluctuations

in sire evaluation. J. Dairy Sci. 62:760-767.

Cooper, J. B. and Hargrove, G. L. 1982. Age and month of calving adjustment

of Holstein protein, milk and fat lactation yields.

J. Dairy Sci. 65:1673-1678.

Cunningham, E.P. 1965. The evaluation of sires from progeny test data.
Anim. Prod. 7:221-231.

Cunningham, E.P. and Henderson, C.R. 1966. Analytical techniques for
incomplete block experiments. Biometrics 22:829-842.

Cunningham, E.P. and Henderson, C.R. 1968. An iterative procedure for
estimating fixed effects and variance components in mixed model
equations. Biometrics 24:13-25.

Dempfle, L. 1977. Comparison of several sire evaluation methods
in dairy cattle breeding. Livest. Prod. Sci. 4:129-139.

Dempfle, L. and Hagger, Ch. 1983. Comparison of the efficiency of BLUP
with other estimstion procedures in dairy sire evaluation.
I. Theoretical investigations. Z. Tierzuchtg. Zuchtgsbiol. 100:196-208.

Dickinson, F.N., Norman, H.D., Powell, R.L., Waite,. L.G. and
McDaniel, B.T. 1976. Procedures used to calculate the USDA-DHIA
Modified Contemporary Comparison. ARS, USDA, Production Research
Report No. 165. p18-34.

Fimland, E.A., Bar-Anan, R. and Harvey, W.R. 1972a. Studies on
dairy records from Israeli-Friesian Cattle. I. Influence of some
environmental effects. Acta Agric. Scand. 22:34-48.

Fimland, E.A., Bar-Anan, R. and Harvey, W.R. 1972b. Studies on
dairy records from Israeli-Friesian cattle. II. Estimates of
components of variance (covariance) and of parameters related to
these components. Acta Agric. Scand. 22:49-62.

Freeman, A.E. and Henderson, C.R. 1959. Genetic structure of
dairy cattle herds in terms of additive and dominance relationship.
J. Dairy Sci. 42-621-625.

Gacula, Jr. M. C., Gaunt, S. N. and Damon, Jr. R. A. 1968. Genetic and
environmental parameters of milk constituents for five breeds.
1. Effect of herd, year, season and age of cow.

J. Dairy Sci. 51:428-437.

Gaunt, S. N., Bartlett, M. A. G. and Comstock, W. L. 1964. Comparison of

five methods of adjusting for environmental differences in evaluation

of dairy sires. J. Dairy Sci. 47:1243-1248.

Hagger, Ch., Schneeberger, M., Crettenand, J. and Schmitz, F. 1984.

Estimation of breeding values for milk yield in Switzerland.

IDF/EAAP Symposium, Prague, September 14-16, 1984.

Harvey, W. R. 1977. Users guide for LSML76. Mixed model least-squares and

maximum likelihood computer program. Ohio State Univ., Columbus

(Mimeograph).

Harville, D. A. and Henderson, C. R. 1967. Estimates of variance components

associated with the first lactations of New York State Holsteins.

J. Dairy Sci. 50:985-986 (Abstr.).

Heidhues, T., Van Vleck, L.D. and Henderson, C.R. 1961. Actual

and expected accuracy of sire proofs under the New York system of

sampling bulls. Z. Tierzuchtg. Zuchtgsbiol. 75:323-330.

Henderson, C.R. 1953. Estimation of variance and covariance components.

Biometrics 9:226-252.

Henderson, C.R. 1956. Cornell research on methods of selecting

dairy sires. Proc. New Zealand Soc. Animal Prod. Conf., 16:69-76.

Henderson, C.R. 1973. Sire evaluation and genetic trends. Proc. Anim.

Breeding and Genetics Symposium in Honor of Dr. J.L. Lush, Am. Soc.

Anim. Sci. and Am. Dairy Sci. Assoc., Champaign, IL, pp 10-41.

Henderson, C.R. 1974. Effect of herd size on accuracy of comparison method

of sire evaluation. J. Dairy Sci. 57:613. (Abstr.)

Henderson, C.R. 1975a. Comparison of alternative sire evaluation methods.

J. Anim. Sci. 41:760-770.

Henderson, C.R. 1975b. Best linear unbiased estimation and prediction

under a selection model. Biometrics 31:423-447.

Henderson, C.R. 1975c. Use of relationship among sires to increase

accuracy of sire evaluation. J. Dairy Sci. 58:1731-1738.

Henderson, C.R., Carter, H.W. and Godfrey, J.T. 1954. Use of the contemporary herd average in appraising progeny tests of dairy bulls. J. Anim. Sci. 13:959 (Abstr).

Hickman, C. G. and Henderson, C. R. 1955. Components of the relationship between levels of production and rate of maturity in dairy cattle. J. Dairy Sci. 38:883-890.

Kennedy, B.W. and Moxley, J.E. 1975. Comparison of genetic group and relationship methods for mixed model sire evaluation. J. Dairy Sci. 58:1507-1514.

Lee, A. J. 1974. Month, year and herd effects on age adjustment of first lactation milk yield. J. Dairy Sci. 57:332-338.

Lee, A.J. and Hickman, C.G. 1972. Age and herd adjustment of first lactation milk yield. J. Dairy Sci. 55:432-438.

Legates, J.E., Verlinden, F.J. and Kendrick, J.F. 1956. Sire and herd interaction in production traits in dairy cattle. J. Dairy Sci. 39:1055-1063.

Lush, J.L. and Straus, F.S. 1942. The heritability of fat production in dairy cattle. J. Dairy Sci. 25:975-982.

Mahamunulu, D.M. 1963. Sampling variances of the estimates of variance components in the unbalanced 3-way nested classification. Ann. Math. Statist. 34:521-527.

Mao, I.L., Burnside, E.B., Wilton, J.W. and Freeman, M.G. 1974. Age-month adjustment of Canadian dairy production records. Can. J. Anim. Sci. 54:533-541.

McClintock, A.E. 1982. Derivation of genetic parameters and environmental adjustments factors for use in the design and implementation of breeding programs for British Friesian cattle. Ph.D. Thesis, Univ. of Reading, Reading, U.K.

Milk Marketing Board. 1984. Dairy Progeny Testing Scheme (Friesians/

Holsteins). Report of the Breeding and Production Organisation of
the Milk Marketing Board of England and Wales. 34:44-45.

Miller, R.H. 1964. Regression of a sire's breeding value on various
functions of daughters' and herd-mate production.
J. Dairy Sci. 47:305-309.

Miller, P.D., Lentz, W.E. and Henderson, C.R. 1970. Joint influence
of month and age of calving on milk yield of Holstein cows in the
Northeastern United States. J. Dairy Sci. 53:351-357.

Miller, R.H., McDaniel, B.T. and Plowman, D. 1968. Comparison of
three methods of sire evaluation. J. Dairy Sc. 51:782-791.

Moriello, F.J. and Legates, J.E. 1970. Genetic differences among
Holstein-Friesian herd improvement registry herds.
J. Dairy Sci. 53:908-912.

Norman, H. D., Kuck, A. L., Cassell, B. G. and Dickinson, F. N. 1978. Effect
of age and month of calving on solids-not-fat and protein yield for
five dairy breeds. J. Dairy Sci. 61:239-245.

Patterson, H.D. and Thompson, R. 1971. Recovery of inter-block
information when block sizes are unequel. Biometrika 58:545-554.

Pirchner, F. and Lush, J.L. 1959. Genetic and environmental
portions of the variation among herds in butterfat production.
J. Dairy Sci. 42:115-122.

Quaas, R.L., Everett, R.W. and McClintock, A.E. 1979. Maternal
Grandsire model for sire evaluation. J. Dairy Sci. 62:1648-1654.

Robertson, A. and Rendel, J. 1954. The performance of heifers got by
artificial insemination. J. Agri. Sci. Camb. 44:184-192.

Robertson, A. and Asker, A.A. 1961. The genetic history and breeding
structure of British-Friesian cattle. Emp. J. Exptl. Agric. 19:113

Robertson, A. and Khishin, S.S. 1958. An analysis of variation
between dairy-sire progeny groups for milk yield and fat content.
Proc. British Soc. Animal Prod., 1958, pp.19-29.

Robertson, A. and McArthur, A.T.G. 1955. Genetic differences between

    bull-breeding herds. Proc. British Soc. Animal Prod. 1955, p.94.

Robertson, A., Waite, R. and White, J.C.D.   1956. Variations in

    the chemical composition of milk with particular reference to the

    solids-not-fat. II. The effect of heredity.

    J. Dairy Res. 23:82-91.

Sanders, H. G. 1927. The variation in milk yields caused by season of the

    year, service, age and dry period and their elimination. .Part I. Season

    of the year. J. Agric. Sci. Camb. 17:339-379.

Sargent, F. D., Butcher, K. R. and Legates, J. E. 1967. Environmental

    influences on milk constituents. J. Dairy Sci. 50: 177-184.

Searle, S.R. 1962. Age and herd effects in New Zealand dairy cow

    records. J. Dairy Sci. 45:82-85.

Searle, S.R. and Henderson, C.R. 1959. Establishing age-correction

    factors related to the level of herd production.

    J. Dairy Sci. 42:824-835.

Syrstad, O. 1965. Studies on dairy herd records. II. Effects of

    age and season of calving. Acta Agric. Scand. 15:31-64.

Thompson, R. 1969. Iterative estimation of variance components   for

    non-orthogonal data. Biometrics 25:767-773.

Thompson, R. 1979. Sire evaluation. Biometrics 35:339-353.

Thomson, G.M. and Freeman, A.E. 1970. Environmental correlations in

    pedigree estimates of breeding value. J. Dairy Sci., 53:1259-1265.

Tucker, W. L. and Legates, J. E. 1962. Effective use of herd-mates in dairy

    sire evaluation. J. Anim. Sci. 21:976 (Abstr.).

Van Vleck, L. D. 1966. Correlations among records of unrelated cows

    in the same and different year-seasons. J. Dairy Sc. 49:61-64.

Van Vleck, L.D., Heidhues, T. and Henderson, C.R. 1961. Analysis

    of deviations of dairy records from different contemporary

    averages. J. Dairy Sci. 44:269-281.

Van Vleck, L.D. and Henderson, C.R. 1961. Ratio factors for adjusting monthly test-day data for age and season of calving and ratio factors for extending part lactation records. J. Dairy Sci. 44:1093-1102.

Van Vleck, L. D., Wadell, L. H. and Henderson, C. R. 1961. Components of variance associated with milk and fat records of artifically sired Holstein daughters. J. Anim. Sci. 20:812-815.

Wilkinson, G.N., Eckert, S.R., Hancock, T.N. and Mayo, O. 1983. Nearest neighbour (NN) analysis of field experiments. J. Roy. Stat. Soc. B 45:151-211.

Wood, P.D.P. 1970. The relationship between the month of calving and milk production. Anim. Prod. 12:253-259.

Wunder, W.W. and McGilliard, L.D. 1967. Season of calving and their interactions with age for lactational milk yield. J. Dairy Sci. 50:986 (Abstr).

APPENDIX

Comparison of Jerseys of different origins for milk and fat yield in India

V.P.S. Chauhan*

Institute of Animal Genetics,
University of Edinburgh,
West Mains Road,
Edinburgh EH9 3JN
Scotland

* on leave from the National Dairy Development Board, Anand 388001, India

purebred Jersey Bull Mother Farms of the National Dairy Development Board, Anand in India.

Introduction

Jersey animals as well as frozen semen have been introduced into the tropics from several temperate countries for dairy cattle breeding. However no information is available on the comparison of performance of genotypes of different origins. Besides the differences in milk, fat and protein yield, it is of great interest in an animal breeding programme to know the rank of a genotype in the temperate and tropical conditions. Although studies from Britain (Robertson *et al.,* 1960), the United States (Legates *et al,* 1956; Van Vleck, 1963 and McDowell *et al,* 1975) and Ireland (Cunningham and O'Byrne, 1977) have shown that the genotype x environment interaction is small, it may be of some interest to examine this problem in the tropical conditions where the performance of exotic breeds is affected by the stress of hot and humid climate as well as the management level.

The aim of this study was to compare the milk and fat yield of Australian, Danish and American Jerseys maintained at the

TABLE 1 - Distribution of animals by herd and genetic group and the number of sires in each group

Herd code	Genetic group						
	AUS	DAN	USA	ENGF	NZF	DANF	INDF
1	77	6	34	2	6	52	12
2	—	94	24	—	4	58	—
3	—	6	—	—	—	—	—
4	—	21	26	8	6	—	39
5	—	13	7	—	—	—	10
6	4	—	21	1	3	7	26
7	—	69	—	3	2	—	23
8	—	29	24	—	—	—	13
9	—	—	32	—	—	10	1
Total	81	238	168	14	21	127	124
No. of sires	49	70	72	5	4	18	36

TABLE 2 - Least squares analysis of variance of first lactation milk yield, fat% and fat yield

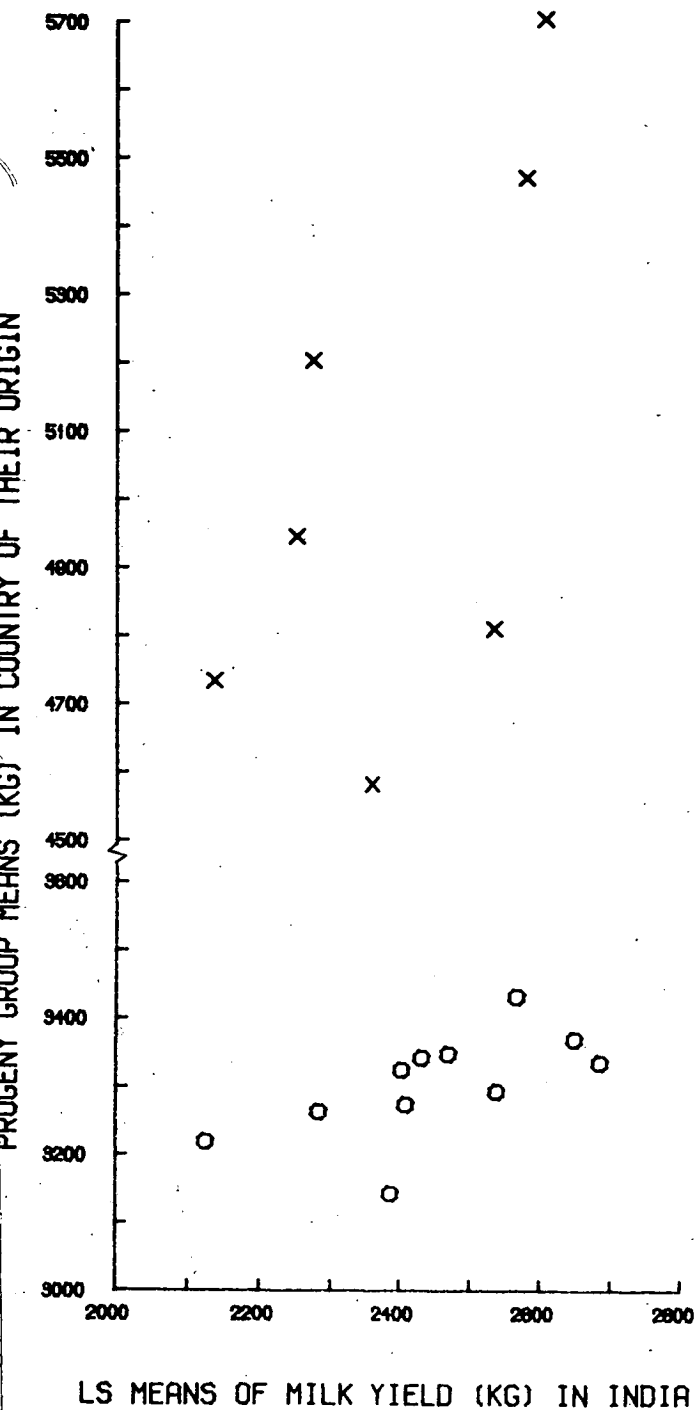| Source	Milk yield		Fat%		Fat yield
	D.F.	Mean squares	D.F.	Mean squares	Mean squares
Genetic group	6	1088108**	5	0.4856*	4135.59**
Sire within genetic group	247	265354	135	0.1693**	980.16
Herd	8	5920915**	5	1.5947**	5882.39**
Year of calving	10	418995	8	0.3750**	2601.88**
Month of calving	11	753564**	11	0.3049**	2136.08**
Regression on age at calving:					
Linear	1	2144025**	1	0.0031	3345.98*
Quad.	1	331884	1	0.0003	256.56
Cubic	1	1122795*	1	0.1433	1214.94
Residual	487	268014	247	0.1162	850.56

* P < 0.05; ** P < 0.01

TABLE 3 - Least squares means of first lactation milk yield, fat% and fat yield

Genetic group	Milk yield (kg)		Fat%		Fat yield (kg)
	No. of obs.	Mean ± S.E.	No. of obs.	Mean±S.E.	Mean±S.E.
AUS	81	2333±203.7			
DAN	238	2557± 75.3	137	5.52±0.11	164.06± 7.56
USA	168	2450± 91.9	98	5.26±0.09	144.84± 6.89
ENGF	14	1978±163.6	8	5.19±0.18	118.77±13.43
NZF	21	2333±139.2	12	5.30±0.16	142.77±11.27
DANF	127	2294± 83.0	100	5.27±0.11	141.04± 6.94
INDF	124	2251± 77.9	60	5.14±0.08	132.27± 5.94
Overall	773	2314± 51.6	415	5.28±0.08	140.63± 5.67

Materials and Methods

The first lactation records of Jersey heifers imported from Australia in 1971-72, from Denmark in 1973 and 1977, from the United States in 1979 and of their farm born heifer herdmates were used to compare different genetic groups. The imported animals were highly selected on pedigree and most of them were born to proven sires. Animals of each country were designated as a separate genetic group: Australian (AUS), Danish (DAN) and American (USA). The farm born heifers were classified according to their sire groups: 5 sires from England and Wales (ENGF), 4 sires from New Zealand (NZF), 18 imported Danish young bulls (DANF) and 36 farm bred young bulls (INDF).

Only records on lactations completed after 250 days were used in the analysis. Shorter lactations were omitted as they were considered to be due to individual health problems. Lactations following abortions or stillbirths were excluded. For lactations continued beyond 305 days the first 305-days

× USA SIRES (ME)
○ DANISH SIRES

LS MEANS OF MILK YIELD (KG) IN INDIA

(Y-axis: PROGENY GROUP MEANS (KG) IN COUNTRY OF THEIR ORIGIN)

However the sire X lactation interaction was not significant.

Finally the following model was used to analyse the first lactation records by least squares (Harvey, 1977):

$$Y_{ijklmn} = U + G_i + S_{ij} + H_k + Y_l + M_m + b_1A + b_2A^2 + b_3A^3 + e_{ijklmn}$$

where,

Y_{ijklmn} = record of nth animal of jth sire of ith genetic group, lactating in kth herd, calved in lth year and mth month

U = overall mean

G_i = fixed effect of ith genetic group

S_{ij} = random effect of jth sire of ith genetic group

H_k = fixed effect of kth herd

Y_l = fixed effect of lth year of calving

M_m = fixed effect of mth month of calving

b_1A, b_2A^2, b_3A^3 = regression on age at calving,

e_{ijklmn} = random error associated with Y_{ijklmn}

The herd-wise distribution of animals in different genetic groups and the number of sires in each group are given in Table 1. More than 50% of the sires had only one progeny. The AUS genetic group had progeny in only two herds while the DAN and USA genetic groups had a fairly good distribution.

Fat % and fat yield records were available only on 415 animals in the first lactation which were also analysed by the above model. No record of fat% was available on AUS genetic group in the first lactation. Genetic groups were compared by t-test using the inverse elements and sires mean squares as the residual variance.

Results

The first lactation records were analysed using the given model and also by ignoring the effect of sire. The sum of squares of the herd and year were reduced by about 50% by fitting sire within genetic group which shows confounding of the sires' effect with herd and year. The sum of squares of month of calving was not affected by fitting sires.

The mean squares and LS means of milk, fat% and fat yield are presented in Table 2 and 3. As can be expected for the tropical conditions the LS means of milk and fat yield were less than the averages generally reported from their home country. In a recent study (Buvanendran and Peterson, 1980) the first lactation milk yield of Jerseys was 1281 kg. in Sri Lanka while their half-sibs produced 3334 kg in Denmark.

The effect of genetic group, herd and month of calving was significant for milk, fat% and fat yield. The effect of year and sire was significant only for fat% which seems to be associated with the import of year and the genetic group.

Comparisons based on first lactation records among AUS, DAN and the USA genetic groups showed no significant difference for milk yield but for fat% and fat yield the DAN was significantly higher than the USA genetic group.

Since the AUS genetic group only had sufficient fat% records in the later lactations these were analysed along with all lactations of all the genetic groups assuming homogeneity of variance over lactations as observed from the analysis

ield was used as lactation yield. Lactation length between 250 and 305 days was considered as the characteristic of different genotypes and therefore it was not taken as a covariable in the final model of analysis. However the results were not affected when the lactation length was fitted as a covariable by regression in the model. The milk yield of later lactations was not used in the comparisons as an analysis of 560 records up to third lactation on all the genetic groups showed a significant genetic group X lactation interaction.

29

of 1204 lactation records in these data. The same model was used but with an additional fixed effect for the lactation number. The LS means of fat% and fat yield for AUS, DAN and USA genetic groups based on all lactations were 5.22 ± 0.11, 5.59 ± 0.18 and 5.32 ± 0.12; 125.38 ± 8.95, 152.86 ± 13.75 and 139.03 ± 9.15 kg. respectively. Comparison between DAN and the USA genetic groups based on all lactations showed similar results for fat% and fat yield as that of the first lactation. It was also observed that DAN produced a significantly higher fat% and fat yield than AUS genetic group but the differences between the USA and AUS were not significant.

Discussion

These data were from a project which was not specifically designed to compare the Jerseys of different origins, but the records were quite numerous. From the results of this study it was concluded that there are no significant differences in the milk yield of Australian, American and Danish Jerseys in the Indian conditions. However the Danish Jerseys could be preferred in the situations where fat yield is the main constraint in determining the net returns from dairying.

While comparing the performance of imported animals it was of particular interest to examine whether the genotypes from temperate conditions have the same rank in the tropics. To look into this problem the LS means of milk yield were estimated for 11 sires of DAN and 7 sires of the USA genetic groups with 5 and more progeny. These sires had progeny group sizes of 5 to 28 in India, 49 to 189 in Denmark and 76 to 1750 in the USA. The correlations of predicted breeding values, the R-value for fat yield in Denmark and the Predicted Difference for milk in the USA with the LS means of milk yield (India) were 0.11 and 0.70 respectively. The fat yield records for most of the progeny of these sires were not available in these data. The correlations of progeny group means for milk yield in their home country and in India were 0.63 for DAN and 0.61 for the USA sires. Correlations of progeny group means were converted into genetic correlations by assuming a heritability of 0.25 in Denmark and the USA and 0.20 in India and ignoring the fact that these were selected sires, which were estimated to be 1.04 for Danish and 1.14 for the USA sires. The progeny means are plotted in Figure 1. It was observed that none of the sires with a high progeny mean in their home country had a very low mean in India. These limited data suggest no interaction.

References

Buvanendran, V. & Peterson, P.H.
1980. Genotype-environment interaction in milk production under Sri Lanka and Danish conditions. Acta Agri. Scan. 30:369-372.

Cunningham, E.P. & O'Byrne, T.M.
1977. Genetic correlation of milk production in Britain and Ireland. 28th Ann. Meet. Eur. Assoc. Anim. Prod., Brussels. 22-25 August, 1977.

Harvey, W.R.
1977. Users guide for LSML76. Mixed Model Least-squares and Maximum Likelihood computer program. Ohio State Univ., Columbus, (Mimeograph)

Legates, J.E., Verlinden, F.J. & Kendrick, J.F.
1956. Sire by herd interaction in production traits in dairy cattle. J. Dairy Sci., 39 : 1055-1063.

McDowell, R.E., Wiggans, G.R., Camoens, J.K., Van Vleck, L.D. & St Louis, D.G.
1975. Sire comparisons for Holsteins in Mexico versus the United States and Canada. J. Dairy Sci. 59 : 298-304.

Robertson, A., O'Connor, L.K. & Edwards, J.
1960. Progeny testing of dairy bulls at different management levels. Anim. Prod. 2: 141-152.

Van Vleck, L.D.
1963. Genotype and environment in sire evaluation. J. Dairy Sci., 46:983-987.

Joint effect of age and parity on milk production in Indian conditions

V. P. S. Chauhan
Institute of Animal Genetics, University of Edinburgh
West Mains Road, Edinburgh EH9 3JN, Scotland

Introduction

As the genetic correlation between milk yields at different ages is close to unity (Barker and Robertson 1966; Tong, Kennedy and Moxley 1979; Powell, Norman, and Elliot 1981; Meyer 1984;) the effect of age on milk yield seems to be purely environmental. Since cow evaluation procedures involve large computing needs, most breeding organisations use age adjustment factors to bring the lactation yields to a standardised basis to enable the comparison of animals of different age, and parity is ignored. Lactation number is affected by the breeding performance of the individual animals. Age-yield relationship in animals with a normal and impaired fertility may not be exactly the same. Therefore, it may be useful to consider age and parity together for comparing animals of different parity and age, especially under the situation like India where age correction factors for field use are not available. The differences between the milk yields of cows of different ages but the same parity have been reported to be significant (Buchsteiner 1976, 1978; Skjervold 1978).

This study was undertaken to examine: (i) whether the milk records should be corrected for the effect of age at calving or parity, (ii) what would be gained if corrections were made for the effects of the both.

Material and Methods

The data consisted of 2276 lactation records on 926 Jersey cows in 9 herds of the National Dairy Development Board, Anand in India, over a period of 11 years. Animals were grouped in 9 categories (3 imported and 6 farm bred by source of dam or sire). Lactations following abortions or stillbirths were excluded and only the records completed after 250 days were used. For lactations completed beyond 305 days the first 305 days yield was used as lactation yield. Age at calving was divided into 7 age classes: up to 2.5, 2.5-3.5,...,6.5-7.5 and over 7.5 years. With this classification each parity occurred in at least two consecutive age classes and the total number of age-parity subclasses was 29. The maximum likelihood (ML) procedure suggested by Henderson, Kempthorne, Searle and Von Krosig (1959) to estimate the environmental effects from the records subjected to culling, was used to study the effects of age and parity in these data. Miller, Harvey, Tabler, McDaniel and Corley (1966) also used the same procedure for estimating age effects. The following ML model was used to fit the joint effect of age-parity along with other factors using LSML76 program of Harvey (1977):

$$Y_{ijklmnop} = u + GH_{ij} + C_{ijk} + Y_l + M_m + AP_{no} + e_{ijklmnop}$$

where,

u = overall mean

GH_{ij} = fixed effect of the ith genetic group and the jth herd

C_{ijk} = random effect of the kth cow of the ith genetic group in the jth herd

Y_l = fixed effect of the lth year of calving

M_m = fixed effect of the mth calendar month of calving

AP_{no} = fixed effect of the nth age group and the oth parity

$e_{ijklmnop}$ = random error

The interactions among the effects of herd, year, and month of calving were ignored because the data were limited. The ML analyses of variance shown in Table 1, were set up using four different analysis fitting : (1) age-parity, (2) age ignoring parity, (3) parity ignoring age, and (4) age and parity simultaneously. The effect of age-parity (AP) was replaced accordingly in the above described model. All models were fitted using a repeatability of 0.323 ± 0.025 which was estimated assuming the homogeneity of variance over lactations using a least squares model similar to the one given above but with the effects of age and parity fitted individually. The repeatability in these data was low as compared to some published estimates of repeatability which are close to 0.50 (Maijala and Hanna 1974). It was biased by culling and also perhaps by more environmental variation in these data due to some unaccountable factors such as disease outbreaks, preferential treatment to imported cows etc. The above estimate of repeatability corresponds to a value of 2.096 as the ratio of error variance to the between cow variance added to the diagonal elements of the equations for cows in ML estimation of the fixed effects in the LSML76 program. Substantial changes in the repeatability used had little influence on the effects estimates.

For examining the differences between age groups within the same parity, and between parities in the cows of the same age group only 13 age-parity subclasses were used. The remaining 15 age-parity subclasses were based on only 71 records.

3

Results and Discussion

The sum of squares for the effects of age and parity fitted jointly and individually one after another are given in Table 1. It was observed that the effects of age and parity were mainly confounded with each other. Out of the total variation accounted for ·by the effects of age, parity, and age x parity interaction, 79.2% was accounted for by age alone. In contrast, the effect of parity ignoring age accounted for only 66.3% of the variation, showing that corrections for age were more useful than parity. Fitting age after parity accounted for 21.4% of the variation whereas parity after age accounted for only 8.4% of the variation. The effect of age x parity interaction accounted for 12.3% of the variation and was significant (P<0.05). Mao, Wilton, and Burnside (1974) reported age x parity interaction to be nonsignificant for milk yield but significant for total fat yield. From their results, however, the variation accounted for by age x parity interaction for milk yield was estimated to be 12.5% of the total variation explained by age, parity, and age x parity interaction.

Since the effects of parity after fitting age and age x parity interaction together accounted for a substantial proportion of the variation, the joint corrections would be more useful than correcting for age and parity separately.

Comparisons of ML constants for milk yields at different ages in cows of the same parity and that of different parity in the same age group are given in Table 2. It was observed that the constants at older ages within the same parity were higher than at younger ages. But the differences were significant only up to fourth lactation,

4

which indicated that the animals reached maturity by their fourth lactation at around six years of age. No significant differences were observed between constants for different parities in cows of the same age group except at 5.5-6.5 years, which could be incidental. However, it was noticed that within the same age group the constants for a higher parity were smaller than for a lower parity, for all age groups except 3.5-4.5 years. Buchsteiner (1976) and McClintock (1982, p 41) also reported significant differences between age classes within the same parity.

Summary

The effects of age and parity on milk yield were examined by maximum likelihood (ML) analysis using 2276 lactation records on 926 purebred Jersey cows. Out of the total variation accounted for by age, parity and age x parity interaction, the effect of age ignoring parity accounted for 79.2%, and parity after age accounted for 8.4% only. Whereas fitting the effect of parity ignoring age accounted for 66.3% and age after parity 21.4%. Age x parity interaction was significant and accounted for 12.3% of the variation. These limited data from the Indian conditions suggest that joint corrections for age and parity would be more useful than the effect of age alone.

Acknowledgements

5

Literature

Barker, J.S.F.; Robertson,A., 1966: Genetic and phenotypic parameters
    of first three lactations in Friesian cows. Anim. Prod.,8:221-240.

Buchsteiner, R., 1976: Estimation of age correction factors of the
    Fleckvieh population in Baden-Wurttemberg. Proc. 27th annual meeting
    European Assoc. for Anim. Prod., Zurich, Aug. 23rd-26th, 1976

Buchsteiner, R., 1978: [Effect of age, monthly calving group and calving
    interval on milk yield in Simmentals.] Thesis,Hohenheim Univ.,F.R.G.
    (A.B.A., 48:4509).

Harvey, W.R., 1977: Users guide for LSML76. Mixed model least
    squares and maximum likelihood computer program.Ohio State Univ.,
    Columbus.(Mimeograph)

Henderson, C.R.; Kempthorne, O.; Searle, S.R.; Von Krosig, C.M., 1959:
    The estimation of genetic and environmental trends from records
    subjected to culling. Biometrics, 15:192-218.

Maijala, K.; Hanna, M., 1974: Reliable phenotypic and genetic
    parameters in dairy cattle. Proc. 1st Wld. Congr. Genet. Appl.
    Livest. Prod.,Madrid, session 1, pp541-563.

Mao, I.L.; Wilton, J.W.; Burnside, E.B., 1974: Parity in age
    adjustment for milk and fat yield. J. Dairy Sci., 57:100-104.

McClintock, A.E., 1982: Derivation of genetic parameters and environmental
    adjustment factors for use in the design and implementation of breeding
    programs for British Friesian cattle. Ph.D. Thesis, Univ. of Reading, UK.

Meyer, K., 1984: Estimates of genetic parameters for milk and fat yield
    for the first three lactation in British Friesian cows.
    Anim. Prod. 38:313-322.

Miller, R.H.; Harvey, W.R.; Tabler, K.A.; McDaniel, B.T.; Corley,
    E.L., 1966: Maximum likelihood estimates of age effects.
    J. Dairy Sci.,49:65-73.

Powell, R.L.; Norman, H.D.; Elliot, R.M., 1981: Different lactations for
    estimating genetic merit of dairy cows. J. Dairy Sci ., 64:321-330.

Skjervold, H., 1978: [Effect of age at first calving.] Buskap og Avdratt.,
    30:51-53, (A.B.A., 47:1702).

Tong, A.K.W.; Kennedy, B.W.; Moxley, J.E., 1979: Heritabilities
    and genetic correlations for the first lactations from records
    subject to culling. J. Dairy Sci., 62:1784-1790.

Table 1

Maximum Likelihood analyses of variance of milk yield fitting the effects of age-parity, age ignoring parity, and parity ignoring age

Source	D.F.	Age-parity S.S.	Age ignoring parity S.S	Parity ignoring age S.S.
Age-Parity	28	46 848 840		
Age ignoring Parity	6		37 102 444	
Parity after Age	7		3 962 818	
Parity ignoring age	7			31 068 988
Age after Parity.	6			9 996 274
Age x Parity	15		5 783 578	5 783 578
Residual	2185	458 369 054	458 369 054	458 369 054

N.B. Sum of squares (S.S.) for all the effects are after fitting the effects of genetic group-herd (41 d.f.), year of calving (10 d.f.) and month of calving (11 d.f.)

Table 2

Comparison of parity and age within each other (ML constants [kg] ± S.E.)

Age (yrs)	Parity number 1	2	3	4	5	6	7
< 2.5	-345 ± 66						
2.5-3.5	-92 ± 75	-152 ± 63					
3.5-4.5		106 ± 71	139 ± 63				
4.5-5.5			190 ± 76	83 ± 67			
5.5-6.5@				291 ± 81	88 ± 75		
6.5-7.5					159 ± 99	-11 ± 87	
> 7.5						130 ± 112	41 ± 99
Significance between ages	**		**	**			

** P<0.01;@ Differences between parities within the same age group only significant at 5.5-6.5 yrs.