ESTIMATION OF GENETIC PARAMETERS IN DAIRY CATTLE USING AN ANIMAL MODEL AND IMPLICATIONS FOR GENETIC IMPROVEMENT

Peter Martin Visscher
Ing. Dronten (Holland) 1986
MSc Edinburgh 1988

Ph.D.
University of Edinburgh
1991

Dedicated to the families LAWTON (Rochdale, England) and KOTTELENBERG (Markelo, Holland), for the first lessons in dairy farming and dairy breeding

CONTENTS

ACKNOWLEDGEMENTS

I thank Robin Thompson and Bill Hill for supervising my thesis work. Robin, thanks for so generously giving your time, and for having patience when I was chasing red herrings. If your supply of time and knowledge was a mere exchange for the litres (note the metric) of "proper" coffee you consumed, then so be it. Thanks also for realising that not everything that other people suggest is "dead boring". Bill, thanks for always having some time for me, no matter how busy you were with (administrative) duties, and thanks for entertaining discussions about (the art of) dairy cattle breeding. It was a pleasure working with and learning from both of you.

I thank the Milk Marketing Boards in the U.K. for their financial support which funded this study and made it possible to attend some interesting scientific meetings. I thank Gordon Swanson of the Milk Marketing Board of England and Wales, my third supervisor, for showing continuous interest in my work and for useful comments on several manuscripts.

Thanks to Sue Brotherstone for preparing data files and for lots of assistance in computer software and hardware matters. I thank her roommate, Peter Keightley, for allowing me to use his Sun (don't worry about the pageturns!). I thank Anne Johnson and Jackie Bogie for doing all those 5 minute administrative jobs.

Steve Bishop has made valuable comments on a paper based on chapter 2. I thank Naomi for reading and commenting on various parts of the thesis.

On a personal note I wish to thank:

- My parents, for love and encouragement and for always supporting the sometimes strange moves I seemed to make.
- Friends back home, for their interest in what I was doing. In particular I thank "the boys" (Dinand, Fred, Han, Jan T., Jan W., Jan Z., Joan, and Theo) for their two visits and for the famous weekends in Holland.

- Sara and Chris, for fun holidays in Scotland and in general for being friends, in particular when their friendship was needed (i.e. for putting up with me).
- My buddie Steve, for sharing good and bad times and for our weekly gourmet feasts together with Valerie, Ailsa and Naomi.
- Naomi, for never doubting me, for endless motivation and, of course, for her love and understanding.

Robin Thompson and Bill Hill made various suggestions concerning the studies in this thesis. In particular, Robin suggested investigating the "semi-balanced" design in chapter 2 and validating the use of quadratic likelihood surfaces in chapter 3. Bill questioned the statistical power of using a likelihood ratio test as is used in chapter 3, and this prompted the work presented in chapter 4. Robin suggested to look at powers for cross-classified designs and proposed an algorithm to estimate parameters for this design (chapter 4). The use of likelihood ratios to estimate variances (chapter 7) was suggested by Bill.

PUBLICATIONS

Papers in press from this thesis are:

- Visscher, P.M. (1991). On the estimation of variances within herd-mean production groups. *J. Dairy Sci.* (~~in press~~). 74: 1987 - 1992

- Visscher, P.M., Thompson, R. and Hill, W.G. (1991). Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livest. Prod. Sci.* (in press).

Published papers (in appendix):

- Visscher, P.M. (1990). REML estimates of parameters for fat yield in pedigree herds in the U.K. using an individual animal model; Individual herd analyses vs. a combined analysis. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.* **XIV**:229-232.

- Visscher, P.M. and Thompson, R. (1990). REML estimates of parameters for fat yield in pedigree herds in the U.K. using an individual animal model; Male and female heritability estimates. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.* **XIV**:233-236.

ABSTRACT

The aim of this study was to estimate genetic parameters in the U.K. pedigree Holstein-Friesian (HF) population using an animal model (AM), and to investigate some implications of the results for genetic improvement.

In a theoretical study it was shown that little bias in estimating variances components is introduced by grouping herds according to their mean (milk) production, a common practice for investigating heterogeneity of variance in dairy populations.

For each of 26 large pedigree herds, comprising of a total of 7720 HF cows, variances and h^2 for first lactation fat yield were estimated with residual maximum likelihood (REML) using an AM. The mean fat yield was 212 kg. The mean and range of individual herd h^2 estimates were 0.38 and 0.03-0.80 respectively, and the average standard error of the h^2 estimates was 0.19. Using likelihood ratio (LR) tests it was found that individual herd h^2 were not significantly different from each other ($P>0.05$), but that phenotypic variances differed substantially among herds ($P<0.01$). An investigation into the statistical power of a LR test for small samples showed that it is difficult to detect real differences in individual h^2 if the standard errors of the estimates are relatively large.

Using production records in lactations 1-3 from 100 large Holstein-Friesian pedigree herds, parameters for milk, fat and protein yield in lactations 1-3 were estimated with REML using an AM. The number of records for each lactation was approximately 39000, 26000 and 17000 for lactation 1, 2 and 3 respectively. Heritabilities for the three yield traits were similar: approximately 0.36 in lactation 1 and 0.30 in lactations 2 and 3. Genetic correlations between yield traits in lactations 1 and 2, for example between milk production in first and second lactations, were approximately 0.86. Genetic correlations between yield traits in lactations 2 and 3 were nearly unity. Genetic correlations between yield traits within lactations ranged from 0.58, for milk and fat yield in lactation 3, to 0.91, for milk and protein yield in lactation 1. Genetic

correlations between yield traits between lactations ranged from 0.55, for milk yield in lactation 1 and fat yield in lactation 2, to 0.85, for milk yield in lactation 2 and protein yield in lactation 3. Environmental correlations between traits within lactations were approximately 0.95, and approximately 0.40 across lactations.

The effect of simplifying covariance structures for milk, fat and protein yield in lactations 1-3 on accuracy of selection for lifetime yield was investigated using selection index theory. It was found that applying a transformation to make the traits in lactation 1 independent at the phenotypic and genetic level of to the yield traits in later lactations, and assuming that three new uncorrelated variates were formed, was highly efficient in terms of accuracy of selection when compared to the accuracy of a general multivariate model. This transformation was recommended for a national BLUP evaluation, since it may take account of selection to a larger extent than when performing separate analyses for milk, fat and protein yield.

INTRODUCTION

One definition of the aim of dairy cattle breeding is to increase the economic efficiency of dairy farming by breeding (more) profitable cows. An important parameter determining profitability of dairy cattle is (efficiency of) the production per cow, and data used for studies in this thesis are restricted to milk production traits. To achieve the aim of higher yielding cows through breeding, animals of superior genetic merit for milk production traits should be identified and chosen as parents for the next generation. The traditional method of "identifying" cows and bulls of high genetic merit is to model the biology underlying the expression of production traits and to make predictions about future performances of animals and their progeny using this model. In animal breeding, where often only phenotypic observations are available, the model to describe the observations is usually presented as a statistical model which is based on an underlying genetical model. Recently, a so-called animal model (AM) has become the genetical-statistical model for predicting breeding values in livestock species. The main feature of this animal model is that all relationships between animals with records are taken into account when predicting breeding values.

The aim of the work presented in this thesis was to estimate genetic parameters in the U.K. pedigree Holstein-Friesian (HF) population using an AM, and to investigate implications of results for genetic improvement.

In chapter 1 some problems concerning the above modelling process are discussed, in particular the statistical and genetical assumptions underlying the prediction and estimation of genetic merit in dairy cattle. Chapter 2 deals with potential consequences of parameter estimation in dairy cattle when herds are grouped according to their mean production. In chapter 3 heterogeneity of variance between individual pedigree herds is investigated and chapter 4 is a study about the statistical power of detecting different variances between herds or herd groups. Variance components for milk, fat and protein yield in lactations 1-3, which are required for predicting breeding values, are presented in chapter 5, followed by a study about the

consequences of these parameter estimates for prediction of breeding values in practice (chapter 6). In the final chapter some open questions which arose from discussions in chapters 1-6 are addressed.

CHAPTER 1

SOME THEORETICAL AND PRACTICAL PROBLEMS ASSOCIATED WITH PREDICTION AND ESTIMATION IN DAIRY CATTLE USING AN ANIMAL MODEL

1.1 Introduction

At present, the common method in most countries for evaluating dairy sires is Best Linear Unbiased Prediction or BLUP (Interbull, 1988). Cow evaluation is usually performed separately, using a selection index type procedure (for example, Hill and Swanson, 1983), or a within-herd BLUP evaluation. Increased computer power, faster algorithms and computational shortcuts such as the simple method to construct the inverse of the numerator relationship matrix (Henderson, 1976; Quaas, 1976) now make it feasible to evaluate sires and cows jointly, using a so-called Animal Model (AM). Various countries are in the testing phase or have started to use AM analyses (see e.g. Wiggans *et al.*, 1988a and 1988b; Ducrocq *et al.*, 1990; Jones and Goddard, 1990). Whether using a sire model, a sire-maternal- grandsire model or an animal model for evaluation of dairy cattle data, the effectiveness of using the prediction of the random effects (breeding values) for the achievement of the breeding goal depends on the extent to which the assumptions of the models are violated. What are these assumptions, and what are the consequences if some are clearly not valid?

In this chapter models currently in use for dairy cattle breeding value prediction are discussed in relation to their implicit assumptions. The aim of the chapter is to highlight existing problems in sire and cow evaluation, and to a lesser extent in estimation of variance components, and to discuss possible strategies to deal with these problems, in particular with reference to prediction of breeding values with an AM. If not specified otherwise, referenced data analyses apply to the black-and-white (Holstein-Friesian) population.

1.2 BLUP - Assumptions

When describing the assumptions underlying the evaluation using BLUP a distinction can be made between the general, theoretical assumptions and the assumptions typically made for practical dairy evaluation.

1.2.1 The general statistical model

Consider the linear model: $\mathbf{y} = \mathbf{Xb} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}$; where:

$\mathbf{y}$ = a vector of observations

$\mathbf{b}$ = vector of fixed effects and covariables

$\mathbf{u}_i$ = random effect i, for example a genetic effect (e.g. sire or animal) and environmental effects (e.g. permanent environment, common environment)

$\mathbf{e}$ = random environmental effect

$\mathbf{X}$ and $\mathbf{Z}_i$ are incidence matrices, and the (co)variance matrices of the random effects are:

$$v(\mathbf{u}_i) = \mathbf{A}_i \sigma^2_{u_i} \quad , \quad v(\mathbf{e}) = \mathbf{R}\sigma^2_e \quad , \quad \text{cov}(\mathbf{u}_i, \mathbf{e}') = \mathbf{0}$$

$\mathbf{A}_i$ is the numerator relationship matrix for (additive) genetic effects, and is usually an identity matrix for other random effects. The covariance matrix for residuals, $\mathbf{R}$, is often assumed to be an identity matrix. The assumptions to obtain unbiased predictions of the random effects with minimal prediction error are (e.g. Henderson, 1973):

1) Using the correct (linear) model to describe the data. Departures from this assumption include incorrect preadjustments for certain fixed effects and covariables which do not appear in the model (e.g. age and month of calving, lactation length), and

a non-linear relationship between the observations and the effects in the model. Henderson (1975b) discusses the impact of the use of some incorrect models on estimates and predictions.

2) Variances and covariances are known, or known to proportionality. Although the true covariance structure is never known, departures from this assumption will still yield unbiased, although not "best", predictors of the random effects in an unselected population if the estimates of variances and covariances are obtained through Maximum Likelihood estimation procedures (Kackar and Harville, 1981). "Unbiased" here is defined as $E(u|\hat{u})=\hat{u}$, where the expectation is over the distribution of "true" breeding values given a particular predicted breeding value. Thus the regression of breeding value on predicted breeding value is unity. The more stringent statistical definition of unbiasedness, i.e. $E(\hat{u}|u)=u$, where now the expectation is over the distribution of predicted breeding values for a particular true (unobserved) breeding value, only holds for the trivial case when $cov(u,\hat{u})=v(u)$, i.e. when $\hat{u}=u$. (Incidentally, in the derivation of BLUP $E(u)=E(\hat{u})$ is used to force the predictors to be unbiased, but these expectations are both zero and do not give information about individual true or predicted breeding values.) Estimates of fixed effects and genetic/environmental trends may be biased when estimates of variances and covariances are used. Unfortunately, the bias on true selection response has hardly been investigated. Sorensen and Kennedy (1984) and Sorensen (1989) give some results from simulation, but a prediction of the bias is not presented. This prediction would be of interest to breeding organisations in particular. In chapter 6 some results concerning bias in estimated response are presented.

Homogeneity of variance or normality is not a necessary assumption for BLUP-evaluation if the covariance structure is known, as shown by Gianola (1986) for the case of heterogeneity of variance. For an *unselected* population it is therefore sufficient to know (or use an estimate of) the second moments of the distribution.

1.2.2 The practical statistical model

When applying statistical methods to model biological processes, simplifications are made to approximate the true, usually unknown, underlying biological model. Hence further assumptions are introduced in the practical evaluation models:

Normality: Under the assumption of (multivariate) normal distributions of the random effects, BLUP has the further desirable properties that the earlier assumptions hold for a selected population (Henderson, 1975a) and that it maximises the response to truncation selection. Selection decisions should not depend on the fixed effects, or, more formally, selection should be on a translation invariant function of the observations (Gianola *et al.*, 1988; Henderson, 1990; Fernando and Gianola, 1990), and the data on which selection is based should be included in the analysis. It is not clear what the best strategy is if selection is not based on a translation invariant function of the data. Henderson (1975a) proposed to adjust the BLUP equations to take account of this type of selection, but his conditional selection model is somewhat controversial (Thompson, 1979; Gianola *et al.*, 1988; Goddard, 1990). An example of selection which is not "within levels of fixed effects" (Henderson, 1973) is selection on (group effect + breeding value), which is commonly practiced in dairy cattle breeding. More research is needed to find methods to optimise genetic progress under this type of selection (Gianola *et al.*, 1988; Fernando and Gianola, 1990; Henderson, 1990). If selection is based on some trait which has zero genetic and environmental correlation with the trait being analysed, then the selection bias will (of course) be zero. Meyer and Thompson (1984) discuss the implications of selection on a correlated trait on variance component estimation when the observations of the trait under selection are not included in the analysis. A further discussion on biases in variance component estimation due to culling is found in chapter 5. The assumed linear relationships between the $\mathbf{u}_i$ and $\mathbf{y}$, which follow from normality, is questioned by Dempfle and Grundl (1988), who state that it may be appropriate to test

whether it is true that $\mathbf{y}$ and $\mathbf{u}_i$ follow a bivariate normal distribution.

Homogeneity of variances: The additive genetic effects and the residual error component usually are assumed to be normally distributed with homogeneous variances across levels of fixed and random effects. Departures from this assumption are obvious. A common observation is, for example, that the (residual) variance depends on the mean to some extent. An investigation into heterogeneity of genetic and error variance between individual herds is presented in chapter 3.

Usually the covariance between $\mathbf{u}$ and $\mathbf{e}$ is assumed to be zero. However, there are cases for which the two effects are correlated. Falconer (1983) argues that for dairy cattle, where cows are usually fed according to yield, the generated covariance should be included with the genetic variance, because the environment is thought to be a consequence of the breeding value. Strictly speaking treating this specific environment as "part of the genotype" is only justified if the regression of the environment (i.e. feed in the example) on the true breeding value is constant for all breeding values and environments , which is unlikely to be true. In section 1.3.2 it is proposed that a GxE (genotype by environment) correlation may partly cause the generally observed heterogeneity of variance. Unfortunately this environmental effect is difficult to classify in commercial herds, and Falconer's suggestion seems the most practical.

The covariance between fixed and random genetic effects is usually assumed to be zero, hence it is assumed there is no genotype-environment interaction. Possible departures are a sire by sex interaction (e.g. in beef cattle), a sire by herd interaction (dairy cattle) and in general a genotype by management interaction. See section 1.4 for a further discussion.

1.2.3 The genetic model

Sorensen and Kennedy (1984) and Kennedy *et al.* (1988) describe the genetic properties of an AM. Assuming a normally distributed random genetic effect and using BLUP in a selected population implies the assumption of the infinitessimal genetic model: an infinite number of independent loci each with an infinitely small additive effect. Bulmer (1980) summarised the assumptions and properties of the infinitessimal model in detail. Gene frequencies are assumed to be constant over time. Within family genetic variance is then unaffected by selection and the Mendelian sampling effect is independent of the parental breeding values. The between family variance changes by selection, due to gametic phase disequilibrium (Bulmer, 1971). Both between and within family variance are affected by inbreeding.

Turelli and Barton (1990) questioned the mathematical assumptions underlying the genetic model described above (which they term the "Fisher-Bulmer infinitessimal model"), namely that with infinite loci the distribution of breeding values remains normal under selection. They showed, using multilocus population genetics theory, that under most forms of selection the distribution of breeding values is systematically driven away from normality through generation of third and higher order linkage disequilibria. For a purely additive model, however, a normal approximation of the distribution of breeding values should be sufficient to predict short-term response to selection in most cases (Turelli and Barton, 1990).

Inclusion of the covariance matrix of the random effects in the model, the numerator relationship matrix "**A**" for additive animal effects, will account for a decrease in genetic variance due to selection, genetic drift (Sorensen and Kennedy, 1983), inbreeding and assortative mating (Kennedy *et al.*, 1988).

It has been proposed to include dominance and epistatic effects in the practical linear model (e.g. Henderson, 1988). Including a dominance effect in the usual linear model is only justified if all loci are under complete dominance (Bulmer, 1980); incomplete dominance gives departures from the linear model because of a

non-linear regression between the genotypic values of any pair of relatives (e.g. offspring-parent). Including dominance gives interpretation problems for the infinitessimal model, in fact a finite dominance variance and a finite inbreeding depression are incompatible under the infinitessimal model (Robertson and Hill, 1983):

If d is a dominance effect at a locus, and the number of loci approaches infinity, then

Finite dominance variance implies $d^2 \rightarrow 0$
Finite inbreeding depression implies $d \rightarrow 0$

Therefore, if a finite inbreeding depression is assumed, which seems reasonable since this is observed in practice, the dominance variance is zero under an infinitessimal model. "Therefore all dominance variance in an infinitessimal model derives from (linkage) disequilibrium" (Robertson and Hill, 1983). However, Smith and Maki-Tanila (1990) proposed mixed linear models including both finite dominance variance and finite inbreeding depression. They argued that for some particular (peculiar?) infinitessimal models it is feasible to have finite inbreeding depression and dominance variance in the model. The properties of their proposed models, in particular with respect to selection and drift and sensitivity to small changes in gene frequencies are not clear and need further investigation. There is a particular interest in quantifying dominance variance within dairy populations nowadays because of prospects of mass production of genotypes (cloning). At present, the pedigree structure from field data in dairy cattle is not very suitable for estimating dominance variance (few fulls-sibs and confounding of dominance effects with common environmental effects), but the increased use of embryo transfer and the establishment of nucleus herds will result in more "informative" animal comparisons for estimating dominance variance.
If populations in which breeding values are predicted or variance components are estimated are crossbred, such as, for example, most black-and-white dairy populations in Europe, crossbreeding effects can be fitted in the model as fixed effects or covariates. Ignoring crossbreeding effect such as heterosis and recombination loss may

lead to biases in prediction of breeding values and estimation of variance components (Van der Werf and De Boer, 1989a and 1989b).

Fitting an epistatic effect clearly violates the basic genetic model assumption of independent additively acting loci. Furthermore, Griffing (1960) and Bulmer (1980) showed that, under selection, the selection differentials in additive epistatic effects will give a temporary response to selection, which will be reversed when selection ceases. The linear model type approach therefore does not seem suitable for genetic evaluation fitting epistatic effects. Practical problems are the lack of good priors for the variances of these effects, and the requirement of the inverse of the dominance (and epistatic) relationship matrix (see e.g. Schaeffer *et al.*, 1989, and Chang *et al.*, 1989, for computation of the inverse of gametic relationship matrices).

Departures from assumptions based on the infinitessimal genetic model are, for example, non-normal distribution of random effects, gene effects that are not (very) small (e.g. a major gene segregating in the population), the presence of dominance and epistatis effects and heterogeneity of (additive) genetic variance. For practical purposes the key question seems to be how good the necessarily simplified models are in achieving the aim, i.e. to predict and obtain a response to selection.

1.3 Heterogeneity of variance

1.3.1 Variance heterogeneity between herd groups

One of the assumptions usually made in dairy cattle (sire) evaluation is homogeneity of genetic and residual variance across levels of fixed and random effects. There is abundant evidence, however, of heterogeneity of variance across herds and herd-years. Relevant references to a quantification of heterogeneity of variance in dairy cattle and a discussion concerning possible strategies for dealing with heterogeneity are given in chapter 3.

If heterogeneity of variance across herd(groups) is just a scaling

effect, heritabilities will be the same across herds and both residual and genetic variances will be different in herds with different total phenotypic variances. If environmental effects are the sole cause of heterogeneity, perhaps the easiest case to grasp, heritabilities will be lower in high variance herds. Lastly, a relatively higher genetic variance in more variable herds may cause the heritability to be greater in those herds. Although this case is difficult to interpret, apart from different sire selection strategies, it is found in many studies (e.g. Hill *et al.*, 1983; Lofgren *et al.*, 1985; Boldman and Freeman, 1988; Dong and Mao, 1990). The popular explanation for the latter case, that animals are allowed to express their genetic potential better in high producing, more variable herds, has no relevance to the genetic-statistical model (Hill *et al.*, 1983; Vinson, 1987), where a phenotype is represented as a random environmental deviation from the genotype. However, if the simple model is extended to a multivariate model, as for example in the next section, a variable expression of genetic potential may be accounted for. Different genetic variation in different environments (e.g. herds) may be a case of "environmental sensitivity" (Falconer, 1983 and 1990), which is a form of "pseudo-interaction" (Dickerson, 1962) if the ranking of genotypes across herds is the same. In that case the product moment correlation of performances in any two environments is unity, while the intra-class correlation is smaller than unity.

Famula (1989) raised a potential problem concerning parameter estimation in herd-classes which may have been classified on a function of the parameter of interest. As a result, for example, the sire variance in high-mean herds may be lower because in those herds relatively more high merit sires are represented, which may just be a sampling effect. This potential selection effect is investigated in chapter 2.

Nearly all estimation methods for heterogeneity of variance parameters have been ANOVA or Maximum Likelihood-type procedures under a sire model. It is not clear how the parameters may differ using an animal model estimation procedure, but it is well known that use of daughter-dam information may yield different parameters (see

for example Van Vleck, 1986, and references therein). Within-sire components may show heterogeneity because of the environmental component (as usually is assumed) or because of the unaccounted for genetic component. For an evaluation with an AM it is necessary to obtain a quantification of the problem using the same model of estimation. This is investigated in chapter 3.

1.3.2 Heterogeneity of variance as a GxE correlation

Conceptually, heterogeneity of variance may be explained through a genotype by environment correlation. Suppose the true model for an observation Y is:

$$Y = A + C + E \quad , \text{ with}$$

A = breeding value (random)

C = environmental effect correlated with A (random)

E = residual environmental effect (random).

It may be convenient to think of C as the effect of the amount of concentrates fed to cows on their (milk) production. If A and C are correlated and follow a bivariate normal distribution, it follows that:

$C = \beta A + E_c$ where E_c = error about the regression. Hence,

$$Y = (1 + \beta)A + E_c + E = A' + E'$$

$\beta = \rho_{ca}\, \sigma_c / \sigma_a$, the regression of C on A,

and $A' = A + \beta A$ is the "targeted" breeding value, i.e. the predicted breeding value for unit accuracy if effect C is ignored in the prediction procedure.

Now consider the following cases, assuming σ_a is homogeneous:

β = constant; all farmers feed the cows in a similar way, according to their true BV (breeding value). Hence the variance of the "observed" BV (A') is constant across environments and any heterogeneity of variance is caused by differences in the variance of E, the true environmental variance.

ρ_{ca} = constant; farmers feed their cows according to their BV, but the variation in feeding may differ between herds. As a consequence a higher "observed" genetic variance will be found in the herds where σ_c is largest. This would be a case of allowing animals to express their "genetic potential" better, and thus increasing "genetic" variance.

In practice, both ρ_{ca} and σ_c would differ between herds. Hence heterogeneity of "genetic" variance is observed. Unfortunately C usually is unknown and therefore $A'=(1+\beta)A$ should be regarded as the BV, and differences in variances may be corrected for when necessary. Although the assumed model is a simplification, it gives an alternative explanation for the observed higher genetic variances and heritabilities in more variable herds. The model does not require a variable true genetic variation across environments, which would give interpretation problems for the underlying genetic model. In practice, dairy producers usually feed their cows according to their phenotypic production. Although this would complicate the above model, the basic argument that heterogeneity of variance could be explained by some (hypothetical) unobserved correlated variate should still hold.

1.4 Genotype by environment interaction

1.4.1 Violations of the assumption of independent genotypes and environments

In the usual practical prediction and estimation models it is assumed that interactions between the main (genetic) random effect and the remaining fixed/random effects are non-existent. It is therefore assumed that a GxE interaction is zero; the ranking of genotypes (e.g. sires) is the same across environments.

Falconer (1983) describes departures from this assumption. Firstly, the genotype and environment may be dependent through a correlation; this effect was described in section 1.2.2. Secondly the ranking of genotypes may be different in different environments. Thirdly, Falconer (1983 and 1990) describes the case where the genetic variance depends on the environment where it is measured. This effect is called "environmental sensitivity" and is equivalent to the "pseudo-interaction" of Dickerson (1962) if the ranking does not change across environments. Observed GxE variance components may contain a part which is due to true ranking difference across environments and a part which is caused by differences in genetic variances among environments (Robertson, 1959; Dickerson, 1962). The latter part is Dickerson's "pseudo-interaction". Only true ranking differences give rise to a departure from unity genetic correlation between performances in different environments. This genetic interpretation, a multiple trait approach, was first suggested by Falconer (1952).

1.4.2 Sire by herd effect

A special case of the general GxE interaction is the sire by herd (SxH) effect. Conceptually a sire effect can be thought of as a vector of breeding values if a sire is tested in many herds (Dempfle and Grundl, 1988). If a GxE exists and sires are tested in few herds, only a correlated response will be obtained in other environments (Dempfle and Grundl, 1988). Henderson (1973) discusses the inclusion of an environmental effect for half-sibs in the evaluation model and states that it can be regarded as an interaction component. Conceptually and statistically, a SxH effect is equivalent to a common environmental effect of (half) sibs in the same herd (Meyer, 1987). An explanation for the occurance of a SxH effect is not straightforward. It can be due to a "true" GxE effect, in that the ranking of sires across herds may differ. Perhaps the conclusions of breed rank differences for production/efficiency under low or high concentrate diets (see e.g. Oldenbroek, 1988) can be extrapolated to a within-breed situation. A GxE effect may be due to a scale effect of the observations (see the discussion under heterogeneity of

variance). If farmers somehow give different treatment to half-sibs, whether it be intentionally or unintentionally, this will result in an observed GxE effect, although in this case it is better to speak of a common environmental effect. An inappropriate model of analysis may also result in an observed SxH component of variance (see Meyer, 1987, for details).

Biases in early proofs of natural service bulls (compared with their later estimated breeding value from A.I. progeny) were the main cause for inclusion of a SxH effect in the national sire evaluation model in the U.S. (Norman, 1974; Norman *et al.*, 1972). In retrospect these biases were effectively removed by fitting a SxH effect (Norman *et al.*, 1985). Although the original motivation (Natural service vs. A.I. bulls) may have disappeared, there is no reason to exclude the effect again. A present motivation may be the existance of many farmers' syndicates, i.e. groups of few breeders that test bulls/cows in their herds. True GxE interactions, heterogeneity of variance and testing of "syndicate" bulls all are possible present day sources for a common environmental half-sib effect. The animal model used in the U.S. has the effect incorporated (Wiggans *et al.*, 1988a and 1988b). A recent quantification of the effect for the British situation showed that approximately 3 % of the total variation could be attributed to a SxH effect (Meyer, 1987). The variance due to sire effects in this study was approximately 8 %. Sire by HYS effects gave a larger components (3.2-4.2 %) than a sire by herd effect (2.5-3.2 %). A standardisation of the observations to within-HYS phenotypic standard deviation showed lower "c^2" estimates (c^2= variance components for SxH effect as fraction of the total phenotypic variance), indicating that the observed interaction effect was due partly to scale effect of the genetic variance across herds/herd-year-seasons.

Meyer (1987) gave an interesting illustration of the effect of a SxH component on BLUP sire evaluation. If a constant total phenotypic variance is assumed, whether there is a SxH effect or not, then the amount of information on their sire contributed by n daughters in the same SxH subclass is equivalent to $m = n/(1 + (n-1)k)$ daughters in m sire-herd subclasses, where $k = c^2/(1 - h^2/4)$. For a few values of c^2 and n, the values of m are shown in table 1.1. This table is from

Meyer (1987) and is slightly extended.

Table 1.1: Number of daughters (m) in m subclasses equivalent to n daughters in one subclass for $h^2 = 0.25$.

	c^2		
n	0.01	0.05	0.10
2	1.98	1.90	1.81
10	9.12	6.76	5.10
20	16.63	9.93	6.61
∞	93.75	18.75	9.38

The table shows the effect of including a SxH component in the evaluation, in particular the relative decrease in information for increasing n and c^2. For $n \rightarrow \infty$, m approaches $1/k$. In the U.S. animal model $c^2 = 0.14$ is used (Wiggans *et al.*, 1988b), which with a heritability of 0.25 results in an upper limit of 6.7 "effective" half-sibs per herd. Clearly, this procedure is powerful in restricting extreme values for sires that are represented in few herd-classes. The importance of including a SxH effect is likely to be increased under an AM evalutation, where half-sibs and full-sibs influence each other's breeding values. Both sire by herd and sire by herd-year-season have conceptual disadvantages if different interaction effects are assumed to be uncorrelated (as in the model used by Meyer, 1987). In that case a sire by herd effect assumes a common half-sib effect, regardless of time, while a SxHYS effect assumes half-sibs in different herd-year-seasons are uncorrelated. However, it is not necessary to assume the effects are uncorrelated. By treating HYS as random, Chauhan (1987a) showed that some interblock information can be recovered. Incorporating a covariance matrix for the random HYS effect gives additional information, albeit small (Chauhan and Thompson, 1986). The same strategy could be applied to SxHYS effects, in a way that half-sibs in "neighbouring" HYS are more correlated than half-sibs in the same herd, but calving further apart. Wade *et al.* (1990) used a model assuming that observations had an autoregressive error structure and estimated parameter with REML. The effect of using their model on the accuracy of prediction is expected to be similar to the model used by Chauhan

and Thompson (1986).

1.4.3 Genotype by environment effect in an animal model

In theory it is easy to incorporate a GxE effect in a BLUP animal model: all that is required is some grouping/definition of "genotypes" and "environments" and the covariance structure of the interaction effects, assuming the genotypes and interaction effects are random. At the same time this defines two problems to be considered: Quantitative, what is the order of magnitude of a possible GxE effect and qualitatively, how should genotypes and environments be grouped. A third problem is how to construct an efficient breeding strategy in the presence of a GxE interaction.

Little is known about a GxE within-breed interaction for dairy cattle. Hill *et al.* (1983), using U.K. data, found genetic correlations of sires' performances in herd-groups split according to mean, variance or coefficient of variation (CV) to be close to unity. In similar analyses in the U.S., Carabaño *et al.* (1990) and Dong and Mao (1990) found similar results; their genetic correlation between sire performances in different herd-groups and different states varied from 0.95 to unity. This does not eliminate the possibility of a GxE interaction, since the definition of environments may have been inappropriate to detect such an effect. Intuitively the genetic correlation between performances across environments is unlikely to be unity exactly, hence some interaction may be assumed. The question of course is how much variation can be explained through this effect and whether it is worth the (computational) effort to include it in the BLUP analysis. The sire by herd effect of the previous section can be seen as an upper limit for any GxE component, because that effect is likely to include an environmental (c^2) component. However, it may contain a true GxE component, which will be scaled up by a factor of four. A desirable experiment would be to measure the performances of genetically identical individuals in different environments. Present reproductive techniques such as embryo splitting and cloning could be used to obtain such genotypes.

The most detailed grouping of genotypes and environments would be

individual genotypes by herd. It is not clear how the covariance structure should be defined. It seems logical to assume the covariance matrix within a herd (environment) to be proportional to the **A**-matrix, as proposed by Foulley and Henderson (1989) for a sire model. An alternative strategy is to group genotypes and environments and assume an identity covariance structure. An example is to group the interaction effects as "% Holstein by Geographical District". It may be argued that genotypes (e.g. genetic groups), environments (e.g. production level), and the interaction between the two effects should be treated as fixed. If the groups are carefully chosen this approach may be preferred since it is simple in concept and computation. However, more research is needed to find some suitable grouping strategy and to quantify the interaction component simultaneously.

An interesting problem concerning a GxE effect is which animals should be selected for breeding purposes. Conceptually a breeding value may be represented as a vector of breeding values for all environments (Dempfle and Grundl, 1988). The presence of a GxE interaction will result in different off-diagonal elements (and diagonals if the genetic variation differs between herds) of the covariance matrix of the vector with breeding values. The definition of "breeding value" should therefore be accompanied by the relevant environment for which improvement is desired. For example, the superiority of an animal (genotype) in its own environment may be different from the superiority of its offspring in other environments (Dempfle and Grundl, 1988). For within-herd replacement, i.e. for the same environment, the interaction component could be regarded as a genetic component. In theory we therefore should calculate two (or more) breeding values: one for the environment in which the animal has performed and where its progeny are likely to perform, and one for some average (non-existing) environment. However, this would be a rather impractical situation. In the dairy industry, with many small herds and no obvious environmental grouping, the best strategy may well be to select animals tested over many herds. Progeny testing of bulls, of dams (or dam-families), and in future perhaps of clones, seems the safest way to achieve the fastest genetic progress in the whole population. Selection in any one environment, for example in a

nucleus (MOET) herd, may only give a correlated response in other environments if the interaction component is substantial. Incidentally, the efficiency of any MOET-scheme may heavily depend on the magnitude of the GxE component.

Alternatively, different lines could be selected which are superior for a particular environment (Dickerson, 1962). Again the grouping of environments would be a problem for the dairy situation. Under the quota regulations, an "intensive" vs. "extensive" management (e.g. according to the number of cows per hectare) may provide a grouping strategy, although in this example the breeding aim is likely to be different in both environments.

1.5 Environmental grouping; fixed or random?

Traditionally herd effects or more precisely herd-year-season (HYS) effects have been treated as fixed. Most countries include this effect or more generally a contemporary group (CG) effect, in their sire evaluation model (Interbull, 1988) or animal model. Treating HYS as random would give biased sire proofs if sires were not randomly distributed over HYS-effects, but to overcome this potential problem of selection it is sufficient to treat herds as fixed (Henderson, 1973). However, this has a major disadvantage in the form of loss of information, in particular when herds are small. Small herds or HYS with mainly daughters from one bull would hardly contribute to progeny group comparison in a sire model. Moreover, it is known that cows calving within a short period of time, but in different arbitrary HYS, are likely to have more in common than cows calving at the beginning and end of the same HYS. With small herd sizes the Prediction Error Variance (PEV) of sires can be reduced substantially by fitting herds as random. From a genetic progress point of view it is interesting to ask if small bias should be allowed in order to improve accuracy (see Gianola *et al.*, 1988, for a discussion).

A possible strategy to recover some of the interblock information is treating some environmental effects as fixed and some as random. Chauhan (1987a) tested various models with some environmental effects random and concluded that a model with herds as fixed and periods

(e.g. years) and herd-period-seasons random was "best", in the sense that PEV were smaller and the product moment correlation of sire proofs between two random subsets was close to unity. However, because there may be a genetic or environmental trend over time within herds, a model treating herd-periods (HP) as fixed and herd-period-seasons as random may be "safer" (Chauhan, 1987a). The latter model, with years as periods and months as seasons, was found to be 37 % more efficient than a corresponding model with herd-year-month (HYM) fixed. Chauhan and Thompson (1986) constructed a "rolling months model", in which months were random and a covariance structure of those effects was fitted. It was concluded that for practical purposes a general random months model (with Identity covariance matrix) would be sufficient (Chauhan and Thompson, 1986). Chauhan (1987b) calculated intra-class correlations for cows calving in the same HP or herd-period-season (HPS), in a model where herds were fixed and the HP and HPS effects random. For fat yield the correlations were 0.16 (same HY) and 0.25 (same HYM). The analysis was extended to multiple lactations by Brotherstone *et al.* (1989), who proposed an evaluation model with HP fixed, lactation-herd effects random, months random and lactation-month-herd effects as random. For pairs of lactations the intra-class correlation of cows calving in the same herd-year-lactation class was 0.4-0.45 for log-fat yield. Wade *et al.* (1990) proposed a time-series model to take account of the correlated error structure of observations in different environmental groups (e.g. months).

It is surprising how little attention the strategy of treating some environmental effects as random has had outside the U.K. It would be easy to adapt existing sire evaluation programs to make some environmental effects random, and it may be particularly useful in a multiple lactation animal model. In the U.S. animal model, fixed seasons are flexible from 2 to 12 months, depending on the number of records in a so-called management group (Wiggans *et al.*, 1988a and 1988b). At present, the average number of records per fixed management group for the Holstein-Friesian population is approximately 4 and is increased to approximately 7 by merging different adjoining month of calving groups (G. Wiggans, personal communication). Clearly a substantial improvement can be made by

redefining fixed management groups and treating seasons as random effects.

1.6 Genetic grouping

A usual assumption in the BLUP evaluation is that the random effects are from a normal distribution with zero mean. If this assumption is violated, e.g. when selection has occurred and the information on which selection decisions were made is not included in the data, biased estimates will be obtained, and the genetic variance will be inflated (for directional selection). An example would be the simultaneous evaluation of proven and young, unproven bulls, whilst assuming an identity covariance matrix between the sire effects. Before the inclusion of the relationship matrix in sire evaluation, some grouping was needed to take account of genetic/environmental trends. Sires would be grouped according to year of birth, year of A.I.-stud entry, percentage genes from another breed, or a combination of those (see Interbull, 1988). When it became computationally feasible to include the relationship matrix between sires into the model, due to Henderson (1976), the need for grouping was reduced. In fact, genetic grouping has been a controversial subject ever since. Thompson (1979) showed there are different ways to include group effects in the model, which will yield different predicted sire effects. He also stressed the potential problem of "misgrouping", in particular when sires have few daughters. Fernando and Gianola (1990) give an example of a sire and grouping selection problem, where selection on a biased estimate gave highest genetic progress.

The introduction of the AM for national sire and cow evaluation further reduces the need for genetic grouping, because of the inclusion of all known relationships. According to Henderson (1988) there is now consensus among animal breeders that the only need to include groups is to account for unknown parents. The grouping strategy likely to be used in practice is based on developments by Westell *et al.* (1988), using so-called "phantom parents" assigned to unknown parent groups (if parents are unknown). The basic idea for this strategy was first proposed by Thompson (1979) for a sire model.

In Europe the grouping of black-and-white dairy cattle is likely to remain partly based on the percentage Holstein-Friesian. It is not clear what the efficiency is of a simple regression on HF% in an animal model, as fitted in the analyses in chapters 3 and 5, compared to the more detailed grouping strategy.

1.7 Individual herd data

Dairy cattle herds differ in management and as a result in genetic and environmental parameters such as variances and heritabilities. In theory it would be best to use within-herd parameters to estimate breeding values and "production abilities" (= breeding value plus permanent environment) for that particular herd. In practice the average herd size, at present approximately 100 in the U.K. for milk recorded herds (Swanson, 1991), is usually too small to obtain accurate parameters. Using data from some of the largest pedigree herds in England and Wales resulted in standard errors of individual herd heritability estimates of approximately 0.19 (see chapter 3). Even for large herds, there is the possibility of heterogeneity of variance across years , since many years of data are needed to utilise information of a sufficient number of daughter-dam pairs in the estimation procedure.

Henderson (1973) showed how to evaluate cows from a (closed) herd using all records and relationships, and subsequently showed how to include estimated breeding values from other sources (e.g. national evaluation) in the intra-herd evaluation (Henderson, 1975c). The frequency of a national AM evaluation will be determined by the demand of the industry and the (computer) costs. If computer access/power is not limiting, new data can be incorporated regularly in the evaluation system and breeding values would be available "all year round". If the national evaluation is run, say, twice a year, and the industry requires information in between those evaluations, a within-herd BLUP could be considered. For sire selection a frequency of 2/3 times per year would be sufficient, but farmers need breeding values before making selection/insemination decisions concerning individual cows, therefore preferably as soon as a lactation is ended or early in the next lactation.

Chesnais and Song (1988) present a system for on-farm beef evaluation. Given certain assumptions they propose to use the RHS (Right Hand Side) and diagonal elements for parents from the latest national evaluation, adjust those values for new progeny information and solve the new equations. A decentralised structure of the industry, with for example micro-computers on the farm, would be ideal for this approach (Robinson and Chesnais, 1988). Intra-herd BLUP remains a sub-optimal evaluation, since not all available information is used. A regular, even continuous, updating of the national evaluation set (e.g. each month a few iterates), computational facilities permitted, seems a better strategy.

1.8 Univariate vs. Multivariate analyses

In this section a distinction is made between single vs. multiple trait evaluation for a given lactation and single vs. multiple lactation evaluation for a given trait. This distinction is arbitrary since, for example, if 2 traits are measured in 2 lactations, then a general multivariate analysis with 4 traits could be compared with any model assuming particular covariance structures (see chapter 6).

1.8.1 Single vs. Multiple trait evaluation

Selection for milk production traits is usually on some combination of breeding values for milk, fat and protein. These traits, which are known to be (strongly) correlated, are evaluated separately in most countries for the sire evaluation (Interbull, 1988). Implicitly, this assumes the traits are uncorrelated if BLUP properties are to hold for a population undergoing selection on some function of these traits. Even for cow evaluation, a selection index type procedure in many countries, traits are evaluated independently (e.g. Hill and Swanson, 1983). The question is what can be gained by evaluating the traits simultaneously, taking their correlations into account. In general, there are two possible advantages of multitrait evaluation (see Thompson and Meyer, 1986a, for a review). Firstly, BLUP requires data on which selection decisions were based to be included in the analysis for unbiased predictions of breeding values, so a potential

bias can be reduced or avoided by including traits on which selection may have been based. Secondly, accuracy of prediction is increased by including more information about any particular trait in the analysis. This increase in accuracy results from creating more genetic links in the data and by improving the estimates of fixed effects through better connectedness of the data.

A simple example may illustrate the potential gain in accuracy of a Multiple Trait (MT) evaluation for milk, fat and protein. Suppose the genetic and phenotypic parameters are known, as well as the fixed effects. Then a selection index approach will give the gain in accuracy of a MT evaluation to a univariate evaluation. With the parameters in table 1.2, the relative efficiency of a single trait index over a MT evaluation is shown (in the same table). The parameters are taken from chapter 5. A simple sire model is used for this example, with a variable number of effective daughters recorded for each trait. The relative gain in efficiency (accuracy) would be greater for cows in an AM, because of their lower levels of accuracy. Although this method is an over-simplification of a BLUP evaluation, it shows that some, albeit little, improvement in accuracy can be gained through a MT approach. For this example the gains are relatively low because the genetic and phenotypic regressions of any trait on the other traits are very similar, and therefore other traits contribute little information (Sales and Hill, 1976a). More examples are given in chapter 6. Some countries, e.g. the U.S., Australia and Holland (Interbull, 1988) for practical purposes (i.e. equal design matrices) use the same heritability for all traits in their single trait BLUP. Therefore an extra loss in accuracy is expected, albeit small since the heritabilities of milk, fat and protein yield are similar. If selection is applied to some particular traits, the loss in efficiency is higher by using single trait evaluations. Wilmink (1988) found selection bias to be highest for milk production, which confirms the belief that milk production is the likely trait for early culling in the first lactation.

Table 1.2: Ratio of accuracies for univariate and multivariate (selection index) sire evaluation.

Number of progeny	Relative efficiency (in %) of univariate analyses		
	Milk	Fat (kg)	Protein (kg)
1	95.8	98.8	98.9
2	96.7	99.0	99.1
3	97.3	99.2	99.3
4	97.8	99.3	99.4
5	98.1	99.4	99.5
10	99.0	99.7	99.7
25	99.7	99.9	99.9

	Parameters used		
	Milk	Fat	Protein
Milk	0.39	0.85	0.95
Fat	0.75	0.36	0.88
Protein	0.91	0.81	0.36

Heritabilities on diagonals, phenotypic correlations above and genetic correlations below diagonals.

If it is decided to use a MT evaluation, accurate estimates of the correlation matrices between the traits are needed. It may be better to use single trait evaluations if such estimates are not available or if estimated parameters are inaccurate (Sales and Hill, 1976a and b). Computationally, the MT evaluation will be more demanding because of the more complex (correlated) structure of the **A** and **R** matrix. A canonical transformation (Thompson, 1977; Hayes and Hill, 1980; Meyer, 1985; chapter 6) would reduce the evaluation to essentially separate univariate analyses.

1.8.2 Single vs. Multiple lactations

Two questions arise with regards to a single versus multiple lactation evaluation: what accuracy can be gained by using multiple lactations, and if later lactations are used, should they be treated as repeated records or as different traits? The discussion differs from the general single lactation MT discussion in the sense that

other observations are later in time.

The first question relates to the breeding goal and the genetic parameters. Meyer (1983) found an increase in accuracy of 5-6 % when using a multiple lactation sire BLUP evaluation over two single lactation BLUP evaluations in predicting a linear combination of first and second lactation breeding values. The increase in accuracy can be partitioned into a genetic part, due to the increase in genetic information, and a part due to improved connectedness in the data (a better data structure). If the selection criterion is first lactation 305-day production and the genetic correlation between first and subsequent lactation performance is not unity, little improvement will be made in sire evaluation/selection. However, with an animal model later lactations will substantially improve the accuracy on the cow side where selection is still practised after the first lactation.

It therefore seems logical to include multiple lactations in an AM evaluation, regardless of the breeding aim (which incidentally will be some function of multiple lactation economic production). The U.S. animal (repeatability) model includes lactations 1-5 (Wiggans *et al.*, 1988a). The second question remains however: should later lactations be regarded as repeated records or as different (correlated) traits? A recent analysis, which accounted for selection bias due to culling on first lactation records, showed a very high genetic correlation between first and subsequent lactations (Meyer, 1984). This correlation was 0.91 between first and second and first and third, and 0.96 for lactations two and three. Similar results were reported by Beaumont (1988), who found genetic correlations between pairs of lactations to be greater than 0.89, estimated from the first 3 lactations in the Montbeliarde breed. Estimates for U.K. data using an animal model are presented in chapter 5.

Although these results suggest the traits are genetically nearly identical, this should not be the only criterion for deciding on a repeatability model. Improved genetic connectedness may be an important factor in decreasing the PEV of the random effects. Treating some environmental factors as random would reduce the effect

of environmental connectedness. The genetic links remain however. A repeatability model assumes homogeneity of variance across lactations and equal heritabilities across lactations. There is abundant evidence that later lactations show higher environmental variances (see e.g. Brotherstone *et al.*, 1989; Brotherstone and Hill, 1986; Hill *et al.*, 1983, and chapter 5, for U.K. data). Scaling the data according to within lactation variance (Hill, 1984) or a log-transformation may partly overcome the problem of heterogeneity of variance across lactations. In France (Bonaiti and Boichard, 1990) and Australia (Jones and Goddard, 1990) second and third lactations are weighted with factors of approximately 0.8. This approach assumes that the heritability is lower in later lactations, which seems to be justified when the literature is considered (Maijala and Hanna, 1974). Ignoring the difference in variance and heritability between first and later lactations will result in the information from later lactations being over-emphasised. A repeatability model further assumes that fixed effects are the same for all parities, which has been found to be incorrect (Meyer, 1983 and references therein). However, different fixed effects for different parities can be incorporated in the model if necessary.

It is not fully known what implications use of a repeatability model (compared to using a MT-model) with an AM will have on genetic progress. The tendency of decreasing the sire-offspring generation interval and the relatively low selection intensity for the cow-cow pathway may suggest that the effect will not be very large. Implications of approximating the "true" covariance structures for using simplified models are discussed in chapter 6.

1.9 Conclusions

Some problems associated with dairy cattle prediction and estimation have been tackled, but only a few. For example, the potential problem of preferential treatment and the desirable computing strategy for prediction and estimation have not been discussed. Another problem which requires further research is how to obtain good approximations to the accuracy of the predictions. All changes in evaluation should be tested against the aim of the evaluation and the assumptions

underlying the model of choice. It seems not desirable to violate important assumptions regarding the covariance structure of the observations in order to make it feasible to solve equations from larger data sets. More research is needed on BLUP under selection, in particular how to obtain the largest genetic progress (Fernando and Gianola, 1990; Gianola *et al.*, 1988).

In the following chapters some of the problems discussed above have been investigated. Chapter 2 investigates the possible bias in parameter estimates, as proposed by Famula (1989), when herds are grouped according to their mean production. In chapter 3 genetic and environmental variances are estimated for individual herds and heterogeneity of variance between herds is investigated. Statistical power of likelihood ratio tests as used in chapter 3 is investigated by simulation in chapter 4. Parameter estimates for milk, fat and protein yield in lactations 1-3 are presented in chapter 5, with a discussion about possible practical models approximating the "true" covariances structure of the observations in chapter 6.

CHAPTER 2

ON THE ESTIMATION OF VARIANCES WITHIN HERD-MEAN PRODUCTION GROUPS

2.1 Introduction

One of the assumptions usually made by users of Best Linear Unbiased Prediction (BLUP) evaluation is homogeneity of variance across fixed effect levels (see chapter 1). There is abundant evidence, however, of heterogeneity of variance across herds or herd-year-seasons for milk production traits (see e.g. Boldman and Freeman, 1988; Brotherstone and Hill, 1986; Dong and Mao, 1990; Hill *et al.*, 1983; Lofgren *et al.*, 1985; Mirande and Van Vleck, 1985, and Short *et al.*, 1990, for some recent analyses). Some of the above authors have found a relationship between herd-mean and within-herd (genetic) variance. Typically for those studies, herds were classified according to their mean (milk) production and parameters were estimated within (and between) herd-mean production groups, using a sire model.

Famula (1989) argued that stratifying herds in this way can be regarded as a form of "selection" on sire progeny groups; herd-means may be higher because of the sires represented in those herds, resulting in herd production groups with a selected sample of sires. A "pseudo-heterogeneity" of variance could therefore be induced by selecting herds on their mean production (Famula, 1989). Short *et al.* (1990) supported Famula's caution on the interpretation of parameter estimates when stratifying herds in production groups. However, the results from the simulation study presented in Famula's paper are not clear, because the observed biases in estimated sire variances were probably not significant (standard errors were not presented, but these can be estimated from the presented ranges and the number of replicates). Furthermore, one would expect the "selection" effect to be symmetrical about the overall mean, i.e., a bias in estimating variances from the highest herd-mean group should be similar to a bias from the lowest herd-mean group. This was not observed.

The aim of this study was to qualify and quantify the magnitude of

the above selection effect.

2.2 Methods and results

For various balanced and "semi-balanced" designs, the effect of selection of herd production groups on the estimates of genetic and residual variances can be quantified.

2.2.1 Balanced nested designs of sires within herds

The reduction in the between progeny groups variance depends on the regression of sire progeny mean on herd-mean. This reduction is largest for a nested design of sires within herds, in the absence of herd effects and other fixed effects, because then selection on herd-means is highly correlated with (direct) selection of sires.

Notation:

- h = number of herds in selected group
- n = number of sires per herd
- p = number of progeny per sire
- Y = sire progeny mean
- H = herd mean
- subscript $_s$ = selected
- i= mean of selected group (= selection intensity)

Normality of random effects is assumed throughout this study. Without loss of generality, let the total phenotypic variance in the base population be unity. Then:

$$v(Y) = (1 - t)/p + t \qquad [2.1]$$

$v(H) = \mathrm{cov}(Y,H) = (1-t)/(np) + t/n$, where t is the intra-class correlation in the base population. Then,

$$b_{(Y,H)} = 1; \qquad r^2_{(Y,H)} = 1/n$$

Using simple linear regression:

$$v(Y_s) = (1 - kr^2)v(Y), \qquad [2.2]$$

with k being the reduction in variance for the selected group. (= $i(i-x)$) for truncation selection, where x is the deviation of the truncation point from the mean in standard deviation units).

For a balanced design the orthogonal Sums of Squares for herds, sires and residual from the Analysis of Variance can be equated to their expectations. It can be shown easily that the expectation for the residual variance is the error variance of the base population. The expectation of the Sire Sum of Squares (SSS) is, on conditioning on the herd-mean:

$$E(SSS|H) = E[\ \{\ \Sigma\ p(Y - H)^2\ \}\ |H\ ]$$

$$= pn[\ E(Y^2|H) - E(H^2|H)\ ]$$

$$= pn[\ (1 - r^2)v(Y) + b^2H^2 - H^2\ ]$$

$$= p(n - 1)v(Y), \text{ since } r^2 = 1/n \text{ and } b^2 = 1$$

Therefore the SSS is not dependent on the herd-value; in whatever way the herds are selected, the within-herd SS for sires is unbiased through that selection. The estimated variances in any selected group are therefore unbiased estimators of the population parameters.

Although the expectation of the sums of squares between sire progeny group means is unaffected by selection, the expectation of the variance between the unobserved sire effects is not. The reduction in genetic variance for the selected group can be predicted using the regression of sire values on herd-means. It follows that:

$$E[v(s)_s] = (1 - kr^2/n)v(s), \qquad [2.3]$$

where $r^2 = p/(p + \lambda)$, $\lambda = (1 - t)/t$ and

$v(s)$= sire variance in the base population.

2.2.2 Selection on progeny means with overall mean as only fixed effect/Ignoring herd-effects

For the limited case of 1 sire/herd (n=1; h=total number of sires), i.e. ignoring herds and selecting solely on progeny means, it can be shown that the expected estimated sire variance in the selected group is:

$$E[\ \sigma_s^2] = v(s)\ [\ 1 - k(\lambda + p)/p\ ] \qquad [2.4]$$

The expectations in [2.3] and [2.4] are identical only in the case of no selection, i.e. k=0, or for the trivial case of λ= 0. The term between the square brackets can become negative for $k > p/(p + \lambda)$, that is, if the repeatability of the predicted sire effect is smaller than the reduction is variance for the selected group. If the ordered progeny group means are divided into four groups by symmetric truncation about the mean, then the largest reduction in estimated sire variance is expected in the two middle groups, because the variation between progeny means is the smallest in those groups. If each group contains exactly 25% of the population, then it can be shown that the reduction in variance for the two middle groups is 0.95 (= k). Of course the distribution of progeny means is symmetric, so that selection of the top or the bottom groups should yield identical results. Famula (1989) used fixed truncation points to obtain four groups each containing approximately 25% of the herds.

Table 2.1 shows a few combinations of the number of sires in the base population (m), h^2 and p, together with predictions of estimated parameters and simulation results. Records were simulated as a sire effect plus a random error term, and evaluated with an analysis of variance (ANOVA), fitting an overall mean and a between and within sire term. For the examples given the heritabilities were chosen to be large, because for low heritabilities and few daughters per sire (highly) negative estimated sire variances were expected (for repeatability $<<$ k). The number of replicates was chosen to obtain sufficiently small standard errors of the mean estimates, and varied

for different sets of parameters.

Table 2.1: Observed and predicted results for selecting on progeny means

				Observed parameters from simulation			Predicted parameters Using formulae (see text) [2.2]	[2.4]	[2.3]
m	p	h^2	group	$v(Y_s)$	σ^2_s	v(s)	$v(Y_s)$	σ^2_s	v(s)
48	10	1.0	4	7.996 (.064)	0.509 (.066)	10.45 (.067)	8.06	0.56	10.54
100	25	1.0	3	1.148 (.008)	-1.858 (.009)	3.618 (.019)	1.15	-1.85	3.59
100	50	0.50	3	0.590 (.007)	-1.172 (.007)	1.973 (.018)	0.58	-1.17	1.98

Phenotypic variance simulated = 100 (units)^2
m = number of sires in base population
p = number of progeny per sire
standard error of simulation results between parenthesis
groups: 4= top 25 % ; 3= second ("next") 25 %

As expected, the simulation results agree well with the predictions. Although this model, for which the criteria on which selection took place are ignored, is unlikely to be used in practical situations, the results show that even in cases with extreme high heritabilities negative variances may be expected.

2.2.3 Unbiased estimators for balanced designs

For a balanced cross-classified design unbiased estimators of the population variances are again obtained: selection on herd-means now is solely environmental, because the variation between herd-means does not contain a between sire component. Although the between-herd SS are reduced, the expectation of sire and residual SS remains the same. Famula (1989) gave a generalisation for the expected SS for sires using Henderson's Method-3 in his formula 11. It can be shown that the last two terms (the bias) in that formula reduce to zero for balanced designs. In all cases the estimate of the residual variance is unbiased.

Consider the linear model $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$ and $v(\mathbf{y}) = \mathbf{ZAZ}'\sigma_u^2 + \mathbf{I}\sigma_e^2$ with definitions:

$\mathbf{y}, \mathbf{b}, \mathbf{u}$ are vectors of the observations, fixed effects (here HYS) and sire effects respectively, $\mathbf{X}, \mathbf{Z}$ are the known incidence matrices for the fixed and random effects, and $\mathbf{A}$ is the numerator relationship matrix.

Famula (1989) showed, using Henderson's selection model (Henderson, 1975a), the expectation of the reduction in SS for sires after fitting HYS in his formula 11, when selection had been practised on a vector of herd means. This expectation is (the notation has been changed slightly):

$$E_s[\, R(\mathbf{u}|\mathbf{b}) \,] = \text{trace}[\, \mathbf{Z}'\mathbf{MZA} \,]\sigma_u^2 + \text{trace}[\, \mathbf{Z}'\mathbf{MZ}(\mathbf{Z}'\mathbf{MZ})^- \,]\sigma_e^2$$

$$- \text{trace}[\, \mathbf{Q}'\mathbf{ZAZ}'\mathbf{MZAZ}'\mathbf{QH}_0 \,]\sigma_u^2 + (\mathbf{t}'\mathbf{Q}'\mathbf{ZAZ}'\mathbf{MZAZ}'\mathbf{Qt})\sigma_u^2 \qquad [2.5]$$

With,

$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$

$\mathbf{Q} = (\mathbf{P}'\mathbf{X}'\mathbf{XP})^{-1}\ \mathbf{P}'\mathbf{X}'$

$\mathbf{P}$ = a matrix to link HYS to herds.

Matrix $\mathbf{H}_0$ and vector $\mathbf{t}$ depend on the selection process, but are not needed explicitly for the proof.

The first two terms of [2.5] are the standard terms for the unconditional (= no selection) case. The last two terms may result in a bias in the estimated sire variance, since they depend on the unknown $\mathbf{H}_0$ and $\mathbf{t}$. To prove that these terms vanish for balanced designs, it is sufficient to show that the matrix $(\mathbf{X}'\mathbf{ZAZ}'\mathbf{MZAZ}'\mathbf{X})$, which appears in both terms, reduces to a zero matrix.

There are h HYS; each HYS has m and each sire within a HYS has p observations. The vector $\mathbf{y}$ is ordered according to sire within HYS. $\mathbf{J}_i$ is a square matrix of ones of order i and $\mathbf{D}_j$ a block diagonal

matrix with each block a $\mathbf{J}_j$ submatrix. Let the sires be unrelated ($\mathbf{A} = \mathbf{I}$). Then

$$\begin{aligned}
\mathbf{X'ZAZ'MZAZ'X} &= \mathbf{X'ZZ'(I-X(X'X)^{-1}X')ZZ'X} \\
&= \mathbf{X'ZZ'ZZ'X - X'ZZ'X(X'X)^{-1}X'ZZ'X} \\
&= \mathbf{X'D}_p\mathbf{D}_p\mathbf{X} - (1/m)\mathbf{X'D}_p\mathbf{D}_n\mathbf{D}_p\mathbf{X} \\
&= p\mathbf{X'D}_p\mathbf{X} - (p^2/n)\mathbf{X'D}_n\mathbf{X} \\
&= p(np\mathbf{J}_h) - (p^2/n)(n^2\mathbf{J}_h) \\
&= \mathbf{0}
\end{aligned}$$

If sires are related, the data structure becomes "unbalanced" in a sense and the above equation would not necessarily hold. Therefore the assumption $\mathbf{A=I}$ is a requirement for the proof.

2.2.4 Semi-balanced nested designs

A bias does occur, however, for unbalanced designs, because the regression of progeny means on herd means is not constant for all sires. Consider the "semi-balanced" case of n sires nested within herds, with p_{ij} progeny for sire j in herd i. Similarly, b_{ij} is the regression of progeny group mean j on herd mean i. Assume the distribution of progeny numbers over sires within a herd is the same for all herds; for example all herds have (p_1+p_2) progeny records pertaining to two sires, with p_1 and p_2 constant for all herds. Let the sum of all records within a herd be m. Then

$$\begin{aligned}
v(H) &= v\left[\left(\sum_{ij} p_{ij} Y_{ij} \right) \Big/ \left(\sum_{ij} p_{ij} \right) \right] \\
&= (1/\{ \Sigma p_{ij} \}^2) \Sigma[p_{ij}^2 ((1 - t)/p_{ij} + t)] \\
&= (1/m^2) \Sigma[p_{ij} (1 + t(p_{ij} - 1))] \qquad [2.6]
\end{aligned}$$

and

$$\mathrm{cov}(Y_{ij}, H_i) = (p_{ij}/m) v(Y_{ij}) \quad , \qquad [2.7]$$

therefore,

$$b_{ij} = [\, m\, p_{ij}\, v(Y_{ij}) \,]/[\, \Sigma\, p_{ij}\,(1 + t(p_{ij} - 1)) \,] \,. \qquad [2.8]$$

The sire SS, on conditioning on the herd mean, is:

$$E[SSS|H] = E[\, \Sigma\, p_{ij}\, Y_{ij}^2 | H \,] - E[\, \Sigma\, p_{ij}\, H^2 | H \,]$$

$$= \Sigma\, p_{ij}\, v(Y_{ij}) - v(H)\Sigma\, p_{ij}\, b_{ij}^2 + H^2[\, \Sigma p_{ij}\, b_{ij}^2 - \Sigma p_{ij} \,]$$

Now the SSS can depend on the herd-value H. Only for the cases of all $b_{ij} = 1$, i.e. the balanced case, or for the case of $E(H^2) = v(H)$, i.e. $E(H) = 0$, does the formula reduce to the form independent of herd-means.

Averaging over all possible herd values in the selected group gives:

$$E[SSS] = \Sigma p_{ij}\, v(Y_{ij}) - v(H)\Sigma p_{ij} + (i^2 - k)v(H)[\Sigma p_{ij}\, b_{ij}^2 - \Sigma p_{ij}] \qquad [2.9]$$

The first two terms are the usual terms for this design, resulting in an unbiased estimate of the sire variance. The last term is the bias in the SS. The bias for the estimated intra-class correlation is:

$$BIAS(\hat{t}) = \{(i^2 - k)v(H)[\Sigma p_{ij}\, b_{ij}^2 - \Sigma p_{ij}]\}/\{\Sigma p_{ij} - \Sigma p_{ij}^2 / \Sigma p_{ij}\} \qquad [2.10]$$

To illustrate the effect an example is given for $p_{ij} = (1,10)$, $h^2 = 0.25$ and a phenotypic variance of 1.0, in the absence of true herd effects. $p_{ij} = (1,10)$ means that each herd has 11 progeny records, one pertaining to the first sire and 10 to the second sire represented in that herd. Then, using [2.6], [2.7] and [2.8]:

$$v(H) = 0.1374,\ b_1 = 0.66 \text{ and } b_2 = 1.03$$

Selecting the top/bottom 25% of the herds ($i = 1.27$, $k = 0.77$, $i^2\text{-}k = 0.85$) gives the bias in the SS of 1.458 per herd (using [2.9]), and hence a bias in the estimated heritability, from [2.10], of +0.03. Selecting either of the remaining middle groups ($i = 0.32$, $k = 0.95$, $i^2\text{-}k = -0.85$), gives the bias in the heritability of -0.03. These

results were compared with simulation results and were found to agree well. Some more examples are given in table 2.2.

Table 2.2: Predicted biases in heritability estimates from semi-balanced nested design

Sires per herd	Progeny distribution	h^2	Estimated h^2		Bias($\hat{h}^2$)	
			top	middle	top	middle
2	1,10	.25	.282	.218	.032	-.032
2	1,10	.50	.596	.404	.096	-.096
3	1,5,10	.25	.259	.241	.009	-.009
3	1,5,10	.50	.529	.471	.029	-.029
10	1,1,4,4,5,5,6,6,9,9	.25	.251	.249	.001	-.001
10	1,1,4,4,5,5,6,6,9,9	.50	.504	.496	.004	-.004

top = top (or bottom) 25% herds are selected
middle = second (or third) 25% of herds

In extreme cases a substantial bias may occur, but for moderate heritability values and three or more sires per herd, the bias becomes very small.

For the above design the direction of the bias is determined by the sign of the factor $(i^2 - k)$. It follows that the heritability is overestimated from evaluating the top/bottom 25% herds, and underestimated when selecting the "next" 25% groups, the absolute value of the bias being the same for both groups, because the quantity $|i^2 - k|$ is identical for the above groups.

For the limited case of only two sires per herd the result becomes obvious if the covariance between the difference of the two progeny group means and the herd mean is considered. This covariance is:

$$\mathrm{cov}[\ (Y_1 - Y_2), H\ ] = t(p_1 - p_2)/(p_1 + p_2)$$

p_1 = progeny number of sire 1

p_2 = progeny number of sire 2

For the above example this covariance is 0.051, and the regression of the progeny mean difference on the herd mean is 0.37 (for $p_1 > p_2$), which is the difference between the the two regression coefficients. Hence the difference between progeny groups within herds depends on the mean of that herd, although the difference between the sire values remains independent of the herd mean.

2.3 Discussion and conclusion

In practice the regression of progeny mean on herd mean may well be close to zero due to herd-year-season and other fixed effects. Therefore the bias for the estimated parameters and the reduction in true genetic variance in the selected group will both be small. Since young sires usually are distributed over many herds, the "selection" effect is thought to be negligible for most practical evaluations. Famula (1989) simulated 1800 herd-year-season (HYS) effects from 150 herds and 150 sire effects, and randomly assigned 15000 progeny records to (270000) HYS by sire subclasses, resulting in an unbalanced cross-classified design. Regressions of progeny means on herd means were likely to be small, since the expected number of records per sire by herd subclass was 15000/(150x150) = 0.67. Furthermore, the differences between those regression coefficients within any herd were probably small. His results that the higher the mean of the herd-group, the lower the estimated sire variance, can therefore most likely be explained by sampling. In practice there usually is substantial variation within herds due to environmental (e.g. year-season) effects; therefore the regressions of progeny means on herd means are expected to be small. Most likely the "sire selection" effect of stratifying herds on their mean production is therefore negligible. If high producing herds have a different sire selection strategy from low producing herds, inducing an additional covariance between sire and herd values, then heterogeneity of variance is present and will be detected by the estimation methods in use.

CHAPTER 3

ESTIMATION OF GENETIC AND ENVIRONMENTAL VARIANCES FOR FAT YIELD IN INDIVIDUAL HERDS AND AN INVESTIGATION INTO HETEROGENEITY OF VARIANCE BETWEEN HERDS

3.1 Introduction

In dairy cattle the model for breeding value prediction for the 1990s in many counties is, or soon will become, the so-called Animal Model (AM). With the AM cows and bulls are evaluated jointly, using the BLUP (Henderson, 1973) method. In theory BLUP requires the true variances and covariances to be known, but in practice estimates (of the ratio) of the (co)variances are used. Usually the parameters are estimated with a similar model to that used for the genetic evaluation, using a REML (Restricted Maximum Likelihood; Patterson and Thompson, 1971) type estimation procedure. It therefore seems logical to estimate the parameters required for the AM-BLUP using a REML procedure fitting the same Animal Model.

Unfortunately AM-REML algorithms are computationally very demanding, so that estimation of population parameters has to be carried out with relatively small samples. For dairy cattle, one suggestion is to use data from (groups of) individual herds to estimate the population parameters (Swalve and Van Vleck, 1987; Van Vleck and Dong, 1988; Van Vleck *et al.*, 1988). This assures that information additional to paternal half-sib comparisons, for example daughter-dam comparisons, is used, since most daughter-dam pairs are in the same herd. Furthermore, use of individual herd data offers a framework to investigate heterogeneity of variance between herds.

One of the assumptions made by most users of Best Linear Unbiased Prediction (BLUP) evaluation is homogeneity of variance across fixed effect levels. There is abundant evidence, however, of heterogeneity of variance across herds or herd-year-seasons for milk production traits (see e.g. Hill *et al.*, 1983; Lofgren *et al.*, 1985; Brotherstone and Hill, 1986; Mirande and Van Vleck, 1985; Boldman and

Freeman, 1988 and 1990; Dong and Mao, 1990; Short *et al.*, 1990, for some recent analyses). Ignoring heterogeneity of variance has consequences for selection and response to selection. Assuming equal heritabilities between groups, Hill (1984) showed the proportion of animals that would be selected from the more variable herds under mass selection. Vinson (1987) used those results to calculate a loss in response to selection. The theoretically correct proportion to be selected from the more variable groups depends on the heritability and phenotypic variance within each group (Van Vleck, 1988a). For sire evaluation the loss in efficiency is likely to be small if sires are tested across many herd-variance groups (Vinson, 1987). Random testing of bulls is clearly not the case for so-called syndicate sires or for proven sires whose semen is imported into another country. Since conversion of breeding values is based on the predicted breeding values of sires in the latter category (Interbull, 1986), these linear regressions may be biased if expensive semen is used in the more variable herds. If it is not known whether the genetic variance, the environmental variance, or both variances are heterogeneous, the effect on accuracy of selection is not predictable. Using an AM, the effect of heterogeneity of variance on estimated breeding values (EBVs) is unknown.

The aims of this study were to estimate genetic and environmental variances for fat yield in individual pedigree herds using an AM, and to investigate heterogeneity of variance between herds. This is the first time an AM has been used to assess heterogeneity of variance between herds, previous attempts being based on sire models. In order to make appropriate significant tests for the estimates, likelihood ratio (LR) tests were used. This involved validating approximations of likelihood functions.

3.2 Material

Production records from the Milk Marketing Board of a sample of 26 large Holstein Friesian (HF) pedigree herds, selected on the number of heifers present in 1986, were taken. After editing, 7720 first lactation fat yield records were present from cows calving between 1981 and 1986. Some summary statistics for individual herds are

presented in table 3.1: 574 sires were represented in the complete data set, both young and old (proven) sires; 186 sires had only 1 daughter, whereas proven sires had up to 450 daughters present; 1740 daughter-dam pairs with records were present, of which only 6 pairs were not in the same herd.

Table 3.1: Summary statistics for individual herd parameters of fat yield.

PARAMETER	MEAN	MIN	MAX	Q1	Q3	STDEV
Mean (kg)	212.4	170.3	263.6	189.9	228.1	26.85
Raw σ_p^2 (kg^2)	1247.1	625.0	2391.2	967.5	1532.8	411.1
No. records	296.9	168	485			
No. animal effects	500.1	329	841			
$r^2(\sigma_p^2, mean)$= 0.59						

The statistics are respectively: mean, minimum, maximum, lower quartile upper quartile and the empirical standard deviation.
Raw σ_p^2= phenotypic variance before any corrections.
$r^2(\sigma_p^2, mean)$= empirical correlation between herd means and herd phenotypic standard deviations.

3.3 Methods

The following linear model was fitted:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \quad \text{and}$$
$$v(\mathbf{y}) = \mathbf{ZAZ}'\sigma_a^2 + \mathbf{I}\sigma_e^2 = \mathbf{ZGZ}' + \mathbf{R} \quad ; \text{ with the usual definitions}$$

y, **b**, **u** are vectors of the observations, fixed effects and individual animal effects respectively, **X**, **Z** are the known incidence matrices for the fixed and random effects, and **A** is the numerator relationship matrix. Herd-year-seasons (HYS) were the only fixed effects, and age at calving, percentage North American Holstein Friesian and lactation length were fitted as covariables. Three seasons of four months were defined as December-March, April-July and August-November, which correspond to the season definition for the current U.K. sire evaluation. Years were defined as from August to July. All sires were treated as "base" animals, hence relationships between sires were not

fitted, in part because many sires had ancestors from foreign populations (the average North American HF percentage of the cows was 23 %). All animal effects, including those of proven sires, were treated as random.

The (natural) Log-Likelihood (L) for a model with one other random effect besides the residual component is (e.g. Harville, 1977; Searle, 1979):

$$L = -\tfrac{1}{2} \{ \log|\mathbf{R}| + \log|\mathbf{A}| + \log|\mathbf{C}| - \log|\mathbf{X'X}| + \mathbf{y'Py} \}$$

where **C** is a full rank submatrix of the coefficient matrix (the matrix containing the left hand side of the Mixed Model Equations [Henderson, 1973]) and **y'Py** is the residual sum of squares, with **P** a projection matrix.

The estimations were carried out using a REML program written by Meyer (1989), which uses an iterative (simplex) search to maximise the likelihood. Consequently, the second differentials (and asymptotic variances) with respect to the parameters are not a by-product of the algorithm. Asymptotic variances of the parameter estimates were calculated by approximating the likelihood surface by a quadratic function in the parameters of interest. This was done by fitting a small grid around the ML estimates. Heritabilities were spaced at intervals of 0.01, and the variances were fitted 1.0 units (kg^2) apart. The matrix of second differentials then gives the realised (observed) Information matrix (see e.g. Fisher, 1956), and its inverse is the asymptotic covariance matrix of the parameter estimates. In the one-dimensional case the approximation reduces to a simple quadratic curve and the second differential matrix reduces to a scalar. The quadratic approximation may also be used within the grid search algorithm. Both these uses of the approximation were suggested by Smith and Graser (1986) for derivative free estimation methods.

Significance tests for heritability and variance estimates were carried out as likelihood ratio (LR) tests (see e.g. Mood *et al.*, 1973), for which 2(L1 - L2) is assumed to follow a Chi-square

distribution if L1 and L2 are the maximum log-likelihoods for different sets of parameters and the parameters in L2 are a subset of those in L1. The quadratic approximation was used to extrapolate the likelihood surface for calculating differences in likelihood for different parameter values. The extrapolation was checked by evaluating the likelihood function at a wide range of parameter values. Likelihood ratios were calculated both for an overall (single) parameter test and for testing individual herd estimates. An overall test (with 25 degrees of freedom) was carried out by calculating an overall estimate for a particular parameter, and comparing the ML pertaining to the overall estimate with the sum of the 26 MLs from the separate herd analyses. The overall estimate was obtained by adding 26 approximated likelihood curves and fitting a quadratic to the newly obtained curve. This approach assumes that parameter estimates from different herds are statistically independent. Individual herd variance estimates were tested in two ways:

1) Assuming the quadratic approximation of the likelihood surface around the maximum, the likelihood for the H_0 (Null-hypothesis) value was maximised and compared with the ML value. This allows the remaining (for the present model only one) parameters to change when comparing the difference in likelihood. For example, if the likelihood surface was parameterised in genetic variance and heritability, then the likelihood was maximised at a value of the genetic variance of 324.5 kg^2, the H_0 value obtained from the combined herd analysis.

2) Differences in likelihood for different variances were calculated at a fixed heritability value. This test is straightforward: using the likelihood equation from above evaluated at a particular heritability value, the likelihoods for different variances are easy to compute. Geometrically, this is looking at a "slice" of the likelihood "mountain" at the fixed heritability value. For this procedure the tests for genetic and environmental variances are equivalent.

Each herd was analysed separately, fitting the above model. To test

the different parameter estimates against some overall H_0 value, a joint herd analysis was carried out, fitting the same model. The estimates from the joint analysis were subsequently used as H_0 values.

Methods to reduce heterogeneity of variance were investigated by using three different transformations of the data. Firstly, data were corrected for the within-HYS phenotypic standard deviation (s.d.). These standard deviations were calculated ignoring other fixed effects and random effects. Data were adjusted in the following way,

$$y^c_{ij} = y_{ij}(sd_p / sd_i), \text{ with } sd_p = \text{population s.d., } sd_i = \text{s.d. for HYS } i$$

and y^c_{ij} = adjusted (transformed) j^{th} observation.

The estimate of the population s.d. was calculated from the ML estimate of the phenotypic variance from the combined herd analysis. An adjustment for HYS s.d. rather than for herd s.d. was made because it is known that within-herd variances are often heterogeneous across years (see e.g. Brotherstone and Hill, 1986), and because HYS rather than herds are usually fitted as fixed effects in the breeding value prediction. Secondly, a (natural) log transformation was made, and finally the square root of the observations were used in the analyses. The latter transformation was made because the log transformation was found to over-correct the data in this study for the mean-variance association.

3.4 Results

The results from the individual herd analyses are presented in table 3.2. Although for all three parameters the estimates were very heterogeneous, only few differed significantly from the overall estimates. The standard errors for the heritability and genetic variance were large, indicating flat likelihood curves. The standard errors for the environmental variances were somewhat smaller, since they were estimated with more degrees of freedom. The average

correlation (not presented) between the genetic and environmental variance estimates for each herd was approximately -0.85. Results from the combined herd analysis are presented in table 3.3. The estimates of the heritability for the complete data set were robust to transformations of the data. The correlation between herd means and estimated herd phenotypic s.d. was 0.71

Table 3.2: Individual herd REML estimates of heritabilities and variances (in kg^2) and results from LR tests.

HERD	h^2	$se(h^2)$	Δ_1	$\hat{\sigma}^2_a$	$se(\hat{\sigma}^2_a)$	Δ_2	$\hat{\sigma}^2_e$	$se(\hat{\sigma}^2_e)$	Δ_3	Δ_4
1	0.33	0.18	0.1	161.8	92.2	3.0	326.3	77.6	7.0*	32.6*
2	0.43	0.16	0.1	224.9	97.5	1.0	299.4	76.1	9.4*	31.8*
3	0.59	0.17	1.6	513.6	178.0	1.3	364.4	130.0	1.7	0.2
4	0.49	0.22	0.3	240.5	122.5	0.5	243.2	95.0	9.3*	33.6*
5	0.03	0.17	4.4!	16.9	3.3	16272.0!	562.4	107.3	0.1	9.6*
6	0.49	0.17	0.4	425.7	168.4	0.4	447.5	132.8	0.4	0.0
7	0.42	0.30	0.0	356.4	274.4	0.0	493.7	222.2	0.0	0.0
8	0.71	0.16	4.5*	433.9	126.1	0.9	184.2	83.9	17.6*	23.2*
9	0.17	0.13	2.6	109.0	63.0	11.8*	449.9	71.0	1.1	20.2*
10	0.28	0.18	0.3	318.5	199.5	0.0	770.4	185.8	1.8	8.8*
11	0.37	0.19	0.0	293.5	161.2	0.0	498.7	137.9	0.1	0.6
12	0.31	0.20	0.1	281.6	179.6	0.0	620.9	156.1	0.4	0.7
13	0.25	0.12	1.1	171.9	80.7	3.4	506.2	78.3	0.1	7.1*
14	0.31	0.18	0.1	239.8	141.0	0.3	528.9	118.5	0.0	1.3
15	0.34	0.19	0.0	318.9	183.1	0.0	615.5	162.1	0.2	1.1
16	0.17	0.12	3.2	174.4	93.9	2.3	819.5	117.4	6.4*	8.1*
17	0.59	0.16	1.7	522.4	178.3	1.3	363.2	129.8	1.7	0.0
18	0.41	0.32	0.0	501.3	422.5	0.2	722.8	328.4	0.3	15.0*
19	0.39	0.23	0.0	352.4	221.3	0.0	551.2	191.8	0.0	0.3
20	0.80	0.20	4.4!	646.0	222.1	2.2	162.0	121.5	11.5*	3.2
21	0.55	0.15	1.3	513.9	168.5	1.3	427.3	123.4	0.7	0.4
22	0.21	0.16	1.1	194.4	132.0	0.8	754.9	147.2	2.3	3.1
23	0.31	0.26	0.1	250.7	205.1	0.1	552.9	188.1	0.0	0.3
24	0.65	0.31	0.8	749.2	415.8	1.2	415.1	323.9	0.1	5.9*
25	0.10	0.11	6.6*	69.7	37.9	63.7!*	580.2	80.1	0.4	5.3*
26	0.38	0.16	0.0	514.7	229.2	0.9	838.8	191.5	2.7	56.0*
COMBINING ESTIMATES (BY ADDING CURVES): HERDS 1-26										
	0.35	0.03	33.7	23.5!	2.4	99.7!*	444.8	22.2	58.4*	268*
SINGLE COMBINED HERD ANALYSIS ESTIMATES, USED AS H_0 VALUES:										
	0.379			324.5			532.3			

$\Delta_{1,2,3}$= -2(difference in log-likelihood) at ML estimate and H0 value
Δ_4= -2(difference log-likelihood) for variances at h2= 0.379
*= significant for P<0.05
!= extrapolation error; estimate is not significant
!*= extrapolation error; estimate is significant at 5% level

A summary of the LR tests is given in table 3.4. The single LR tests showed a significant difference among herds in genetic and environmental variances ($P<0.05$), but not in heritabilities. A single test for the variances at a fixed heritability value of 0.379 (see table 3.3) resulted in highly significant differences in variances ($P<0.01$). The ML variance estimates from adding up the curves were considerably lower in value compared with the estimates from the combined herd analysis. The extreme low value for the genetic variance (23.5) is an extrapolation error; excluding herd 5 from the analysis resulted in an estimate of 180.9 and a LR of 43.2 (still significant).

Table 3.3: Results of combined herd analyses of variances (kg^2) and heritability estimates for fat yield.

ANALYSIS	REML ESTIMATES			
	$\hat{\sigma}_a^2$	$\hat{\sigma}_e^2$	$\hat{h}^2$	$se(\hat{h}^2)$
I Standard	324.5	532.3	0.379	0.037
II Adjustment for HYS σ_p^2	261.3	479.0	0.353	0.036
III Log Transformation	0.0073	0.0123	0.372	0.037
IV Square root transformation	0.378	0.625	0.377	0.037

The likelihood differences in columns 4, 7 and 10 of table 3.2 were from likelihood comparisons with the ML estimates from the combined herd analysis, which were 324.5 kg^2, 532.3 kg^2 and 0.379 for the genetic variance, environmental variance and the heritability respectively (see table 3.3). For two data sets the heritabilities and genetic variances were different from the overall estimate ($P<0.05$). In 6 cases the environmental variance was significantly different from 532.3 kg^2. Assuming the heritabilities to be the same (0.379) in all herds, 13 of the 26 variances were significantly different from the overall estimate (see last column of table 3.2). Therefore, if the heritabilities are assumed to be equal, the phenotypic variance is highly heterogeneous between herds. Testing heritabilities against a H_0 value of close to zero (10^{-4}) resulted in 17 heritabilities differing from that value ($P<0.05$). A single LR

test against "zero" showed a highly significant LR of 205.7 ($P<0.01$, for 25 degrees of freedom).

Table 3.4: Summary Likelihood Ratio tests.

		ANALYSES			
		I	II	III	IV
Number of significant individual herd estimates from 26 separate LR tests					
PARAMETER TESTED					
h^2	P<0.05	2	2	2	2
	P<0.01	0	0	0	1
σ_a^2	P<0.05	2	2	1	1
	P<0.01	2	0	1	1
σ_e^2	P<0.05	6	2	3	4
	P<0.01	5	0	2	3
σ_p^2 |H^2	P<0.05	13	1	8	9
	P<0.01	11	0	8	6
Test statistics from single LR test					
PARAMETER TESTED					
h^2		33.7	32.7	36.9	36.0
σ_a^2		99.7**	87.7**	55.9**	62.3**
σ_e^2		58.4**	23.8	51.6**	44.8**
σ_p^2 |H^2		268.4**	16.9	151.7**	154.7**

Analyses: I = standard, II = data adjusted for within-HYS phenotypic standard deviation, III = Log transformation, IV = Square root transformation.
Values for separate herd LR tests are number of estimates which are significantly different from the H_0 values.
Values for single LR test are -2[difference log-likelihood].
H^2 = H_0 heritability value, taken from combined herd analysis (table 3.3)
! = LR values are overestimates because of extrapolation errors (see text).
σ_p^2 \|H^2 = ML estimate of the phenotypic variance at H^2
** = $P<0.01$

In general, the quadratic approximation overestimated the difference in likelihood between the ML estimates and the H_0 values. In some cases, for example for the genetic variance in herds 5 and 25, this

lead to spurious conclusions regarding the significance of the estimates. The real difference in twice the log-likelihood for these herds was only 3.70 and 5.92 respectively. The curvature at the ML values was much "steeper" than at other points on the likelihood surface.

Adjusting the data for an (uncorrected) estimate of the within-HYS variance resulted in 2 heritabilities, 2 genetic variances and 2 environmental variances (from 4 different herds) being significantly different ($P<0.05$) from the values 0.353, 261.3 and 479.0 respectively, which were the ML estimates for the complete (combined) data set using adjusted records (from table 3.3). Testing the variances at a fixed heritability value of 0.353 resulted in one of the variances differing ($P<0.05$). At the 1% level none of the parameter estimates were different from the overall estimate. A single LR test indicated no significance for all 3 parameters ($P>0.05$).

For the log transformed data, 2 heritabilities, 1 genetic variance and 3 environmental variances for individual herds differed ($P<0.05$) from the H_0 values. However, assuming equal heritabilities (0.372), 8 phenotypic variances were still significant ($P<0.01$), and a single LR test was highly significant ($P<0.01$). The correlation between herd mean and phenotypic variance on the log scale was -0.28. The log transformation slightly "over-adjusted" the data for heterogeneity of variance. The square root transformation, however, showed similar results to the log transformation.

3.5 Discussion

3.5.1 Estimates of individual herd parameters and their implications

Few extreme heritability estimates were obtained despite the relatively large standard errors. The combined herd heritability estimate agrees well with the most recent estimate using a sire model (Meyer, 1987). However, the herds were chosen on size and may not be a representative sample of the pedigree herds, and the complete sample was rather small. Since all sires were treated as uncorrelated

random effects, selection would bias the heritability estimates downwards. Alternatively, an increased variance might be expected as the sires were from different populations (European and North-American).

Apart from two rather high estimates (for herds 8 and 25), the heritabilities were similar. More data per herd would increase the ability to distinguish between different heritability estimates, but the herds were the largest available, and the average herd size in the U.K. is the largest in Europe. If no inference could be drawn from these samples, it is not clear how AM herd estimates should be obtained. A multi-lactation analysis would increase the amount of information substantially, but a multi-trait evaluation is computationally very demanding and may require different computing algorithms (Meyer, 1991). The overall, single, LR test may be more suitable for inferences about the population, since sampling will usually result in some individual estimates different from the mean value.

The results suggest that the heritabilities are relatively constant and that the phenotypic variance is heterogeneous. The crude correction for the heterogeneity of phenotypic variance, by adjusting data for within-HYS phenotypic standard deviation, reduced the heterogeneity substantially. Despite the relatively large correlation between herd mean and herd variance, the log transformation over-adjusted the data for heterogeneity. The resulting negative correlation (-0.28) between herd mean and herd variance indicates that if this transformation is applied in a BLUP analysis, assuming a constant heritability among herds, the breeding values of superior cows from high yielding herds would be underpredicted relative to the breeding values of superior cows from low yielding herds.

Existing literature estimates of heterogeneity of variance are often contradictory both between countries and within countries over time. While some studies find a correlation between herd-mean and herd-(phenotypic)-variance (Mirande and Van Vleck, 1988; Hill *et al.*, 1983; Brotherstone and Hill, 1986; Meinert *et al.*, 1988; Boldman and Freeman, 1988 and 1990), others find no evidence of such a

relationship (Lofgren *et al.*, 1985; Winkelman and Schaeffer, 1988). Even for the studies that did find a (positive) correlation, the relationship was not strong. A typical value would be 0.4-0.5 (for milk, fat and protein yield). Hence heterogeneity of variance cannot be explained fully by a scale effect. With a correlation not very close to unity, the log transformation seems to reverse the trend, in that the association between mean and variance becomes negative.

Previous studies to quantify heterogeneity of variance were often based on grouping herds according to some criterion and estimating variances using a sire model. Grouping on herd-mean (Mirande and Van Vleck, 1985; Boldman and Freeman, 1988 and 1990), herd-variance (Winkelman and Schaeffer, 1988) or on a function of the mean and variance, e.g. the coefficient of variation (Hill *et al.*, 1983; Lofgren *et al.*, 1985; Pearson *et al.*, 1988) are the usual choices. Lofgren *et al.* (1985) found no clear pattern of heritability estimates by grouping herds on herd-mean. The "average" herd-mean class had the lowest heritability for milk yield (0.163). They found consistently higher heritabilities in the more variable groups. The effect of their implicit assumptions, unrelated sires and all sires from the same population, on the obtained estimates is not clear, but the heritability estimates were probably biased downwards. Mirande and Van Vleck (1985) looked at trends in genetic-environmental variances over a 22-year period. Within-sire variances increased over time, thus decreasing the heritability. It is perhaps not surprising that parameters should change over such a time period. The trait itself may well have changed (genetically) in that time, in such a way that the genetic correlation between measurements on the same trait in different time periods is less than unity. It is debatable if the same pre-adjustment factors for certain "fixed" effects can be used for cows calving that far apart. Heritabilities for fat yield were found to be higher in both high-mean and high-variance herd classes (Hill *et al.*, 1983). A log-transformation indicated that the difference in variance was a greater cause of those higher heritabilities than the high herd means. Results from daughter-dam regression within herd classes according to phenotypic standard deviation and herd-mean indicated that heritability estimates for milk yield would be a function of the herd-variance (higher standard

deviation showed higher regression coefficient) and not of the herd mean (Pearson *et al.*, 1988). Boldman and Freeman (1988, 1990) found similar results: the high herd-production groups showed higher genetic and environmental variance and a higher heritability. There seems evidence that the heritability is consistantly higher in the more variable herds. The conclusion concerning the relationship between herd mean and heritability is less clear.

An interesting question is what causes heterogeneity. Possible explanations include management factors (e.g. feeding, housing), breeding strategy (sire selection), genotype by environment interaction, a common environmental effect for half-sibs (i.e. a herd-sire effect) and preferential treatment. For the present analysis, a potential sire-herd effect was confounded with the genetic variance. Similar results regarding heterogeneity of variance may not be expected using an AM compared with using a sire model, since the within-sire component may be heterogeneous because of environmental variance or because of the unaccounted for genetic component.

Usually the aim of estimating parameters is to use them subsequently in, for example, a BLUP evaluation. The question therefore is what strategy should be used to deal with the problem of heterogeneity of variance between environments. Ignoring it altogether is the simplest option, and this may not have been too inefficient until now, when sires and cows are evaluated separately, assuming sires were tested over many herd-variance groups and that heritabilities are higher in the more variable herds. For a separate cow evaluation, the problem of heterogeneity of variance is potentially much more serious: ignoring the effect will have a cumulative effect over time, given a selection index type approach and the fact that most cows will have female ancestors producing in the same herd (Vinson, 1987). The cow genetic index (CGI) in the U.K. standardises observations to the within-HYS phenotypic standard deviation, by regressing the estimate of a within-HYS standard deviation to an overall standard deviation depending on the variance of the estimate (Brotherstone and Hill, 1986). The (national) genetic progress is affected if it is less efficient that more bull-dams come from the more variable herds as

will be the case if correction does not take place. The justification for no correction would be that the heritability is also higher in the more variable herds. With an AM it seems unjustified to ignore the effect, although the effect of heterogeneity of variance on accuracy of selection is not clear. Unfortunately biases are difficult to predict since they depend on the structure of the data and the true parameters. Simulation should indicate what the loss in efficiency is for certain population structures and parameters.

Hill (1984) showed a standardisation to within-group phenotypic standard deviation is justified if the heritability is constant across groups. Meinert *et al.* (1988) found this strategy to give the best results for the regression of daughter on her sire's predicted transmitting ability. For the present data set this correction seems to be sufficient. A disadvantage of this adjustment is that it requires regular estimates of within-herd variances, preferably corrected for fixed effects, if the data are to be precorrected for heterogeneity of variance. For small herds (i.e. most herds), this may give sampling problems. Using a Bayesian argument, parameters from individual herds could be regressed to some overall mean according to their accuracy (sampling variance), as in Brotherstone and Hill (1986). However, the within-herd parameters are likely to change over time. Brotherstone and Hill (1986) found repeatabilities for most parameters (mean and variances) between herd-years to be about 0.7, but even so, changes in management may cause abrupt changes in parameters (Mirande and Van Vleck, 1985); for example, the effect of quota introduction in Europe on (genetic) parameters is unknown. Alternatively, the adjustment could be made in the estimation program. Again, however, sampling effects should be taken into account.

A log-transformation has been proposed and investigated by various authors (e.g. Hill *et al.*, 1983; Meinert *et al.*, 1988; Boldman and Freeman, 1988, 1990), based on the evidence of a correlation between herd mean and variance. The log-transformation is justified if the heterogeneity is just a scale effect, resulting in the standard deviation being linearly related to the mean. If the mean-variance correlation has no genetic component, a log-transformation will have

the additional advantage of increasing the heritability. If the relationship is (partly) genetic, the heritability may be different on a log-scale, depending on what proportion of the mean-variance correlation is genetically determined. Hill *et al.* (1983) found within-sire variances of log-yields stabilised across herds grouped on the mean, but between sire components relatively unaltered. Hence the overall heritability increased and the difference between high and low increased after the log-transformation. For herds split according to variance the ratio of within-between sire components before and after the log-transformation remained fairly constant. Even given the higher heritability in high mean and high variance herds, the weights given to untransformed records from those herds in a sire evaluation were theoretically too large (Hill *et al.*, 1983). Heritabilities for milk yield, for low, medium and high herd-level groups remained nearly constant after a log-transformation, but the low-level group (with the lowest heritability for both untransformed and transformed yield) had the relatively highest phenotypic variance after the transformation (Boldman and Freeman, 1988 and 1990). Superior cows in low producing herds would therefore be overevaluated on the log-scale; unadjusted yields would overevaluate cows from the high-level group. These findings are confirmed in the present study. Caution should therefore be taken in applying a log-transformation, since the genetic and environmental variances may not respond the same way to this transformation. In the present study both variances seemed to respond similarly to the transformation, although the genetic variance was not very heterogeneous to start with. Brotherstone *et al.* (1989) and Brotherstone and Hill (1986) looked at within-sire heterogeneity of variance by adjusting records for the breeding value of the sire, and concluded that a log-transformation would reduce the heterogeneity. Correcting for a daughter's sire, by subtracting her sire's transmitting ability, assumes homogeneity of genetic variance, which is inconsistent with previous studies (Hill *et al.*, 1983). The log-transformation therefore cannot solely be justified by looking at the reduction in heterogeneity.

If further investigation indicates that heritabilities are not the same for all herds, then a different approach should be taken. A multi-trait approach seems theoretically best (see e.g. Schaeffer *et*

al., 1978; Gianola, 1986), but it may be tedious to estimate genetic and phenotypic parameters for all herds in order to group them according to some function of the estimated parameters. Furthermore, grouping herds according to genetic and/or environmental variances would give sampling problems (Winkelman and Schaeffer, 1988).

Given the literature findings and the results from the present study, it seems most practical to pre-adjust data for some estimate of the herd or HYS phenotypic standard deviation.

3.5.2 The use of quadratic approximations in LR tests

A quadratic approximation of the likelihood surface was used to obtain asymptotic (co)variances and to extrapolate the likelihood surface for testing parameters. The latter use gave spurious likelihood differences for variances when the H_0 value to be tested was not close to the ML value. Apparently, although perhaps not surprisingly, the likelihood surface does not "behave" as a quadratic function over a wide range of parameter values. One way to investigate the slope of the likelihood surface is to examine the geometric curvature at different parameter values; for a perfect quadratic surface the curvature, here defined as minus the second differential of the likelihood with respect to the parameter(s) of interest, is constant for all parameter values. The curvature for a particular parameter at the ML estimate is called (Fisher's amount of) Information.

Table 3.5: Curvature of log-likelihood at various values of the genetic and environmental variance for a one-way balanced design.

	ADDITIVE GENETIC VARIANCE (σ_a^2)				
	0.30	0.35	0.40	0.45	0.50
ENVIRONMENTAL VARIANCE (σ_e^2)					
0.50	1092.9	849.9	673.7	541.9	440.8 $\Psi(\sigma_a^2)$
	1068.6	869.1	714.3	592.1	494.1 $\Psi(\sigma_a^2,\sigma_e^2)$
	1305.3	1077.5	895.9	749.4	630.0 $\Psi(\sigma_e^2)$
	-0.89	-0.91	-0.92	-0.93	-0.94 $r(\sigma_a^2,\sigma_e^2)$
0.55	892.8	695.9	552.6	445.1	362.4 $\Psi(\sigma_a^2)$
	846.4	692.1	571.6	475.8	398.5 $\Psi(\sigma_a^2,\sigma_e^2)$
	1022.7	850.2	711.4	598.3	505.3 $\Psi(\sigma_e^2)$
	-0.89	-0.90	-0.91	-0.92	-0.93 $r(\sigma_a^2,\sigma_e^2)$
0.60	737.3	575.4	457.3	368.5	300.1 $\Psi(\sigma_a^2)$
	677.6	556.4	461.2	385.1	323.4 $\Psi(\sigma_a^2,\sigma_e^2)$
	809.6	677.0	569.3	480.8	407.6 $\Psi(\sigma_e^2)$
	-0.88	-0.89	-0.90	-0.91	-0.92 $r(\sigma_a^2,\sigma_e^2)$
0.65	614.5	479.9	381.4	307.2	249.8 $\Psi(\sigma_a^2)$
	547.5	450.9	374.7	313.6	263.7 $\Psi(\sigma_a^2,\sigma_e^2)$
	646.4	543.1	458.5	388.5	330.3 $\Psi(\sigma_e^2)$
	-0.87	-0.88	-0.90	-0.91	-0.92 $r(\sigma_a^2,\sigma_e^2)$
0.70	516.3	403.1	320.1	257.5	209.1 $\Psi(\sigma_a^2)$
	445.9	367.9	306.2	256.5	215.8 $\Psi(\sigma_a^2,\sigma_e^2)$
	519.9	438.3	371.1	315.3	268.4 $\Psi(\sigma_e^2)$
	-0.86	-0.88	-0.89	-0.90	-0.91 $r(\sigma_a^2,\sigma_e^2)$

Ψ = curvature matrix= -[matrix of 2nd differentials]
$r(\sigma_a^2,\sigma_e^2)$ = correlation between estimates derived from the Ψ-matrix
True parameters: $\sigma_a^2 = 0.40$, $\sigma_e^2 = 0.60$

For illustration, following Visscher and Thompson (1990, see appendix), consider a one-way balanced half-sib design, with 100 sires each having 10 recorded offspring. Using true values of the heritability and phenotypic variance of 0.40 and 1.0 respectively, the curvature for different combinations of parameter values for the genetic and environmental variance is presented in table 3.5. Clearly the curvature changes with different parameter values. Visually, this is demonstrated in figure 3.1, which represents likelihood contours for various combinations of the values of genetic

and environmental variances from table 3.5, using both exact likelihoods and likelihood values obtained from a quadratic approximation of the likelihood surface at the ML values. Close to the ML values the quadratic approximation seems sufficient, but departures from a perfect quadratic surface are clearly visible for more extreme values of the variances. A different parameterisation, for example in heritability and phenotypic variance, gave similar results. The magnitude of the extrapolation error is illustrated in table 3.6. For different values of estimated heritabilities, the LR was calculated as twice the difference in log-likelihood and compared with the LR obtained from approximating the likelihood curve by a quadratic around the ML estimate. For this example, the predicted LR overestimated the true difference in log-likelihood when testing values larger than the ML value, and underestimated the difference for values smaller than the ML value. The extrapolation error is rather small for the example given, but this reflects the flat likelihood curve for a heritability estimate based on 100 progeny groups of 10 half-sibs.

Table 3.6: Exact and predicted Likelihood Ratios (LR) for a balanced design.

	H^2(ML)									
	0.20		0.30		0.40		0.50		0.60	
h^2(t)	LR_1	LR_2	LR_1	LR_2	LR_1	LR_2	LR_1	LR_2	LR_1	LR_2
0.20	0.0	0.0	1.4	1.2	5.1	3.8	10.9	7.3	18.4	11.2
0.30	1.2	1.5	0.0	0.0	1.1	1.0	4.2	3.2	8.9	6.3
0.40	4.3	5.9	1.0	1.2	0.0	0.0	0.9	0.8	3.5	2.8
0.50	8.5	13.2	3.6	4.6	0.8	1.0	0.0	0.0	0.8	0.7
0.60	13.5	23.5	7.1	10.4	3.0	3.8	0.7	0.8	0.0	0.0

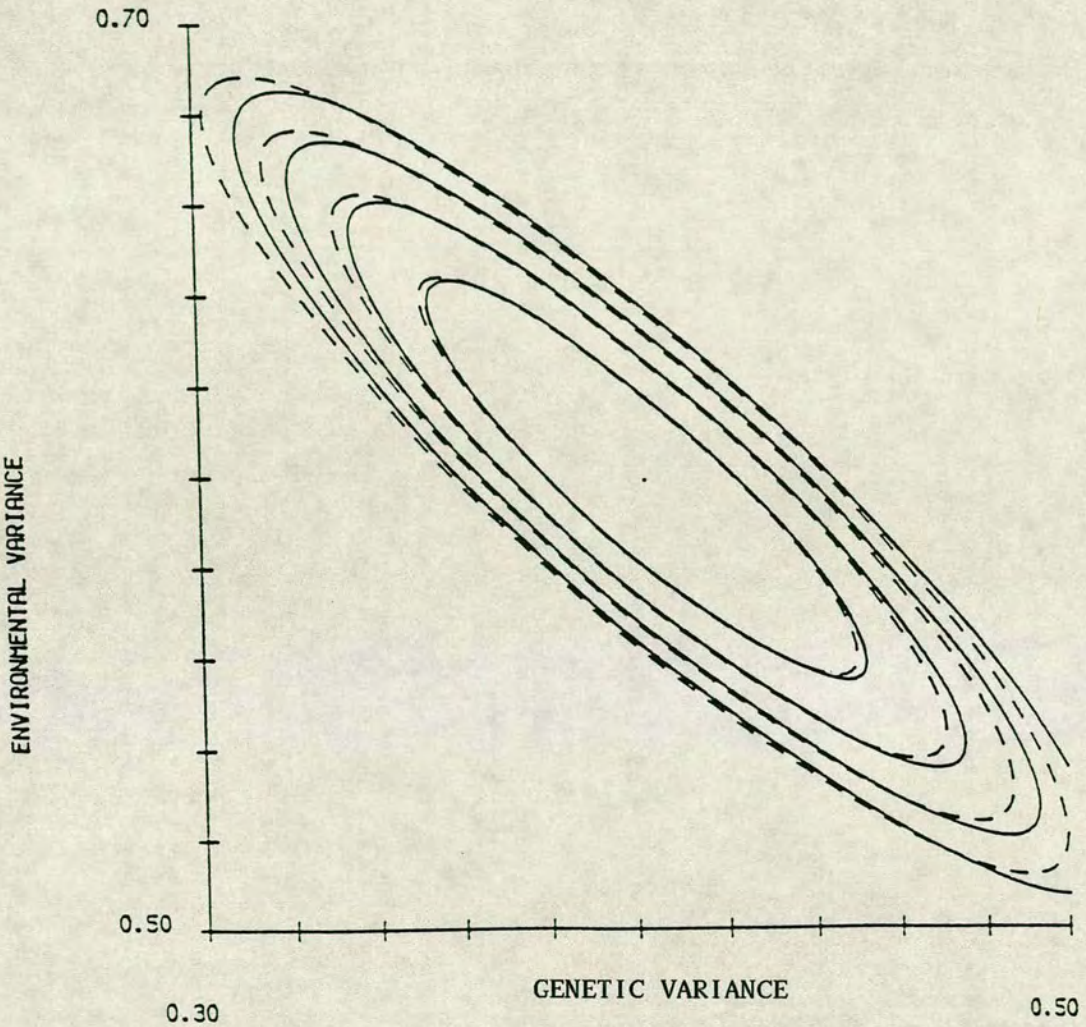
H^2(ML)= Maximum Likelihood estimate
h^2(t)= heritability estimate which is tested against ML value
LR_1= exact LR from likelihood curve
LR_2= predicted LR from quadratic around the maximum

Various authors (e.g. Smith and Graser, 1986; Graser *et al.*, 1987) have suggested use of a quadratic approximation of the likelihood surface to obtain asymptotic variances when the second differentials or the expectations thereof are not a by-product of the estimation

algorithm. However, in data analysis and simulation it has been found that a quadratic approximation sometimes does not produce sensible results, in particular when many random effects are estimated (Meyer, 1989). Visscher and Thompson (1990) discussed differences in curvature at different parameter values for a hierarchical nested design. For the example given here, a one-way balanced design, the argument is analogous: since the variances of the Mean Squares depend on their expected values, and the parameters of interest are linear functions of the Mean Squares, the curvature depends on the values of the parameters. A cubic approximation would produce better results, since the second differentials are still functions of the parameter values, but if there are many random components, for example in a multiple trait analysis, this would require a large multi-dimensional grid and the inversion of a rather large matrix. Using a quadratic approximation for a multi-dimensional grid search may not be efficient, so transformations of the parameters to make the likelihood surface more quadratic may speed up convergence.

In the discussion above it was argued that more data per herd would be needed to increase the ability to distinguish between different individual herd parameter estimates. The relationship between the size of individual herd data sets and the ability to detect differences in variance estimates can be viewed as a problem of statistical power. This is investigated in the next chapter.

Figure 3.1: Likelihood contours for a balanced half-sibs design



Parameters used: $\sigma_e^2 = 0.60$, $\sigma_a^2 = 0.40$ and data on 10 progeny of 100 sires.

Differences between subsequent contour lines is 0.10 log-likelihood

Solid curves: contours for exact likelihoods.

Dashed curves: contours for likelihoods obtained from a quadratic approximation of the likelihood surface around the maximum likelihood values.

CHAPTER 4

ON THE POWER OF LIKELIHOOD RATIO TESTS FOR DETECTING HETEROGENEITY OF INTRA-CLASS CORRELATIONS AND VARIANCES IN BALANCED HALF-SIB DESIGNS

4.1 Introduction

In animal breeding, BLUP has become the method of choice for predicting breeding values from mixed linear models. Theoretically, (co)variances of random effects included in the mixed model should be known without error, but in practice estimates thereof are used. It has become standard practice to estimate variances using REML. The most desirable (linear) model both for prediction of breeding values and estimation of genetic parameters appears to be an (individual) animal model (AM), in which relationships between all animals in the data and pedigree are taken into account (e.g. Wiggans *et al.*, 1988a and 1988b, and Smith and Graser, 1986, for applications in dairy cattle).

One assumption usually made by users of BLUP is homogeneity of variances across levels of fixed (and random) effects. In dairy cattle, however, there is abundant evidence that this assumption is not valid (see e.g. Lofgren *et al.*, 1985; Mirande and Van Vleck, 1985; Brotherstone and Hill, 1986; Boldman and Freeman, 1988 and 1990; Short *et al.*, 1990; Dong and Mao, 1990, for some recent analyses). Typically for studies investigating heterogeneity of variance, herds or herd-year-seasons (HYS) are grouped according to their mean production or phenotypic variance, and parameters are estimated within and between herds or HYS using a sire model. Unfortunately, using an AM for estimating parameters is computationally demanding, and relatively small sample sizes are necessarily used to estimate population parameters. One suggestion for dairy cattle parameter estimation is to use individual herd data as samples (see e.g. Swalve and Van Vleck, 1987; Van Vleck and Dong, 1988; Van Vleck *et al.*, 1988; chapter 3), and to combine several individual herd estimates into a population estimate. Using

individual herd data separately provides a framework to investigate heterogeneity of variance between herds (Van Vleck and Dong, 1988; chapter 3). If results about variance heterogeneity from a sample of individual herd estimates may be extrapolated to the total population (of herds), then, for any trait and parameterisation in heritability (h^2) and phenotypic variance, either one of the following conclusions may be drawn from one such sample:

1) Both heritabilities and phenotypic variances are homogeneous across herds
2) Heritabilities are homogeneous, phenotypic variances are heterogeneous across herds
3) Heritabilities are heterogeneous, phenotypic variances are homogeneous across herds
4) Both heritabilities and phenotypic variances are heterogeneous across herds.

The (arbitrary) parameterisation in heritabilities and phenotypic variances, instead of parameterisation in additive genetic and environmental variances, was chosen to investigate the conclusions from chapter 3 about heterogeneity of variance between herds which were in terms of the same parameterisation. Furthermore, results from estimating variances in dairy cattle are commonly reported in h^2 and phenotypic variances. The implications of these four scenarios for a (national) BLUP evaluation, if the appropriate covariance structure of the data is to be considered, vary substantially. Scenarios 2 to 4 imply that estimates for individual herds should be obtained regularly, which is tedious and may be subject to sampling error. Furthermore, besides (sampling) problems associated with estimation of the relevant parameters, there may be computational problems with a large scale implementation. For example, scenarios 3 and 4 suggest a general multi-trait approach (Gianola, 1986), which may not be feasible for computational reasons, even if all parameters were known.

Inference about the (co)variance structure of observations across herds or HYS therefore has implications for the choice of the desirable model to be used. A question that arises is what significance test should be used in deciding about the most likely

scenario, and how powerful such tests are for small sample sizes. Since the estimation procedure usually is REML, one suggestion is to use a likelihood ratio (LR) test, which has desirable asymptotic properties.

The aim of this study was to investigate the power of a LR test in detecting heterogeneous variances for individual groups (herds). To define and illustrate the problem and to investigate the effect of small samples on departures of test statistics from their expectation (based on large samples), a simple model was used for which sets of group means were tested for equality while allowing for heterogeneous within group variances. Similarly a LR test was used to test whether h^2 differed between herds, while allowing for heterogeneous individual herd phenotypic variances. To predict the power of a LR test for a given design, the distribution of variance estimators are required. Unfortunately, in most practical cases the distribution of AM-REML variance estimates is not known. One suggestion is to investigate the detection of differences in between and within sire variances in different herds using balanced half-sib designs, since the distributions of variance estimators from such a design are known (using ANOVA to estimate variances). Both nested and cross-classified half-sib designs were used to contrast the statistical power in detecting heterogeneous variances across individual herds for these designs.

4.2 Methods

4.2.1 Testing for equal group means with heterogeneous group variances

Suppose there are observations in k groups, and that group means are assumed to be fixed. Assume each group contains n observations and that the observations are normally distributed. Then, for the observed mean in group i (i=1,k),

$$\bar{x}_i \sim N(\mu_i, \sigma_i^2/n) \qquad [4.1]$$

A function of the (unrestricted) log-likelihood (L_u) is, apart from a

constant,

$$-2L_u(x_{ij} \mid \mu_1 \ldots \mu_k, \sigma_1 \ldots \sigma_k) = \sum_i^k \sum_j^n [\ (x_{ij} - \mu_i)^2 / \sigma_i^2\ ] + \sum_i^k n \log(\sigma_i^2)$$

[4.2]

and minus twice the Maximum Likelihood (ML_u) with respect to means μ_i and standard deviations σ_i, is

$$-2ML_u = N + \sum_i^k n \log(s_i^2) \qquad [4.3]$$

With $N = kn$

and $$s_i^2 = \sum_j^n (x_{ij} - \bar{x}_i)^2 / n \qquad [4.4]$$

Now hypothesise that the means μ_i are the same, but allowing for different variances within each subclass. Let the common mean be μ_0, then, setting the first differentials of [4.2] with respect to μ_0 and the σ_i^2 to zero gives:

$$\sum_i^k \sum_j^n [\ (x_{ij} - \mu_0)\ / \sigma_i^2] = 0 \qquad [4.5]$$

And for the variance in group i,

$$\sum_j^n [\ (x_{ij} - \mu_0)^2 / \sigma_i^2] - n = 0 \qquad [4.6]$$

Solving [4.5] and [4.6] requires iteration since there is no explicit solution. Minus twice the ML under the null hypothesis can be written as:

$$-2ML_0 = N + \sum_i^k n \log(s_i^2 + (\bar{x}_i - \hat{\mu}_0)^2) \qquad [4.7]$$

Hence an expression for (a function of) the Likelihood Ratio (= λ), is,

$$\lambda = -2 (ML_0 - ML_u)$$

$$= \sum_i^k n \log[1 + (\bar{x}_i - \hat{\mu}_0)^2 / s_i^2] \qquad [4.8]$$

The LR asymptotically has a ChiSquare distribution with (k-1) degrees of freedom if the null hypothesis is true. The degrees of freedom are from estimating $2k$ parameters for the unrestricted model (see Equation [4.2]) and (k+1) parameters under the H_0 model (see Equations [4.5] and [4.6]). If the means are not the same the distribution of the LR is a Non-central ChiSquare.

For any set of k different group means, the non-centrality parameter is a function of the sum of squares of the fixed means (Kendall and Stuart, 1973, pp. 230-231). Examples of the power of a LR detecting differences in means could be given, but would be conditional on a particular (arbitrary) set of fixed means. To investigate the power of a LR test under the alternative hypothesis (means not the same), one suggestion is to look at an average power from different sets of fixed means. For ease of computation and simulation, and for illustration purposes, sets of means were obtained by sampling them from a normal distribution (to keep the illustration in this section simple, calculations are still based on a fixed effects model, although it could be argued that a random effects model would be more appropriate). Then, if the true group means are repeatedly sampled and for each true mean its estimate is sampled, assuming a multivariate normal distribution of true means and within group observations, it can be shown that the asymptotic distribution of the

test statistic is:

$$\lambda \sim c\, X^2_{(k-1)} \qquad \text{with } c = \left(v(\bar{x}_i | \mu_i) + v(\mu_i) \right) / v(\bar{x}_i | \mu_i)$$

$v(\bar{x}_i | \mu_i)$ is the sampling variance of the estimate of the group mean given the true group mean value, and $v(\mu_i)$ is the population variance of true group means. A full proof of the form of the asymptotic distribution is outside the scope of this study. An approximation of equation [4.8] gives:

$$\lambda \approx \Sigma \left[(\bar{x}_i - \mu_i)^2 / (s_i^2 / n) \right] + \Sigma \left[(\mu_i - \hat{\mu}_0)^2 / (s_i^2/n) \right]$$

$$\approx X^2_{(k-1)} + \left[v(\mu_i) / v(\bar{x}_i | \mu_i) \right] X^2_{(k-1)}$$

$$= \left[\left(v(\bar{x}_i | \mu_i) + v(\mu_i) \right) / v(\bar{x}_i | \mu_i) \right] X^2_{(k-1)}$$

Assuming a joint multivariate normal distribution of true means and their estimates, the (asymptotic) power of the LR test then can be predicted using a central ChiSquare distribution. For significance level α, the predicted power is:

$$P(\alpha) = \int_{\left[X^2_\alpha(k-1) \right]/c}^{\infty} f(x)dx \quad , \quad \text{with } f(x) \text{ the density of a } X^2 \text{ distribution with } (k-1) \text{ degrees of freedom,} \qquad [4.9]$$

c is defined above and $X^2_\alpha(df)$ is the $100(1 - \alpha)$ percentage point for a central ChiSquare distribution with df degrees of freedom.

To investigate the behaviour of the test statistic for small samples, a simulation was carried out using a balanced design. For each replicate, k true means were sampled from a normal distribution with mean μ_0 and variance $v(\mu_i)$. An estimate of μ_i, $\bar{x}_i$, was sampled from $\sim N(\mu_i, \sigma_i^2/n)$ and an estimate of the group variance, s_i^2, was sampled from $\sim (\sigma_i^2/df)X^2_{(n-1)}$, with df being the degrees of freedom (n for ML). The overall mean, μ_0, and the variance within each group were set to 1.0. Therefore, although group variances were individually estimated (using equation [4.4], [4.5] and [4.6]), they were sampled from a homogeneous population (all $\sigma_i = 1.0$). Replicates were varied for different designs to obtain similar standard errors of means over replicates. On average, 10,000 samples were simulated.

4.2.2 Balanced nested half-sib designs

There is an analogy of the previous model to a balanced half-sib design; now consider the groups to be herds (or strata) and a LR test is used to determine whether a particular set of herds differ in intra-class correlation (ICC), phenotypic variance, or in both. The intra-class correlation is the ratio of between sire variance to the sum of between and within sire variance, and is usually assumed to be one quarter of the heritability. One suggestion is to ignore informaton between herds and to assume that individual herd parameter estimates are solely from progeny group comparisons within that herd. Let there be sn observations in each herd, from s sires with n progeny each. Then, assuming normality, the log-likelihood of error contrasts (Patterson and Thompson, 1971) for data from herd i is (see, for example, Thompson and Meyer 1986b), apart from a constant,

$$L = -\tfrac{1}{2} \{ s(n-1)\log(\sigma^2_{wi}) + (s-1)\log(\sigma^2_{wi} + n\sigma^2_{bi})$$

$$+ W_i/(\sigma^2_{wi}) + B_i/(\sigma^2_{wi} + n\sigma^2_{bi}) \}$$

with σ^2_{wi} = within sire variance in herd i,

σ^2_{bi} = between sire variance in herd i,

W_i = within sire SS (sum of squares) for herd i,

B_i = between sire SS for herd i.

Reparameterisation in t_i, the ICC for herd i, and σ^2_i, the phenotypic variance in herd i, gives,

$$L = -\tfrac{1}{2} \{ s(n-1)\log(1-t_i) + (s-1)\log(1 + (n-1)t_i) + W_i/(\sigma^2_i(1-t_i)) + B_i/(\sigma^2_i(1 + (n-1)t_i)) + (sn-1)\log(\sigma^2_i) \}$$

If all herds have sn records, from s sires with n progeny each, and sires are only represented in one herd, then a function of the likelihood for data from k herds, is:

$$-2L_u = \sum_i^k [s(n-1)\log(1-t_i) + (s-1)\log(1 + (n-1)t_i) + W_i/(\sigma^2_i(1-t_i)) + B_i/(\sigma^2_i(1 + (n-1)t_i)) + (sn-1)\log(\sigma^2_i)] \quad [4.10]$$

and the (Residual) Maximum Likelihood is obtained by substituting the ANOVA estimates for t_i and σ_i^2 in [4.10], for $t_i > 0$. Now consider the null hypothesis that the ICC are the same in all herds, whilst allowing for heterogeneous phenotypic variances across herds, and let the common value of the ICC be t_0. Then,

$$-2L_0(x_{ij}|\ t_0,\sigma_1 \ldots \sigma_k) = \sum_{i}^{k} [\ s(n-1)\log(1-t_0) + (s-1)\log(1 + (n-1)t_0)$$

$$+ W_i/(\sigma_i^2(1-t_0)) + B_i/(\sigma_i^2(1 + (n-1)t_0)) + (sn-1)\log(\sigma_i^2)\ ] \quad [4.11]$$

REML estimates for t_0 and σ_i^2 satisfy, respectively:

$$\sum_{i}^{k} \left[\frac{-s(n-1)}{(1-t_0)} + \frac{(s-1)(n-1)}{(1+(n-1)t_0)} + \frac{W_i}{\sigma_i^4\ (1-t_0)^2} + \frac{B_i}{\sigma_i^4\ (1+(n-1)t_0)^2} \right] = 0 \quad [4.12]$$

and $$\frac{W_i}{(1 - t_0)} + \frac{B_i}{(1 + (n-1)t_0)} - (ns- 1)\sigma_i^2 = 0 \quad [4.13]$$

Again, as for the fixed effects example, there is no explicit solution for t_0 and σ_i^2, and iterative techniques must be used to solve [4.12] and [4.13] and to obtain the maximum likelihood estimate. Similar formulas could be derived for the hypothesis that the phenotypic variances are homogeneous while allowing the ICC to differ between herds, or for the hypothesis that both ICC and phenotypic variances are homogeneous.

Using REML, the exact sampling variances of the estimates are not known. One suggestion is to use approximate sampling variances pertaining to ANOVA estimates. Assuming $\hat{t}_i$ is estimated from an ANOVA, its distribution, a non-linear function of a F-distribution, is clearly not normal. However, its sampling variance is known, and for large s and n, $\hat{t}_i$ will be approximately normally distributed. The sampling variance of $\hat{t}_i$ is (from Fisher, 1921), approximately:

$$v(\hat{t}_i) \approx \frac{2[\ 1 + (n-1)t_i\ ]^2\ (1 - t_i)^2\ (sn - 1)}{s(s-1)n^2\ (n-1)} \qquad [4.14]$$

with $E[\hat{t}_i] \simeq t_i$, and s and n, as before, the number of sires and progeny per sire.

If it is assumed that the distribution of the true, unknown, ICC in the population is normal, with variance $v(t_i)$, then, analogous to [4.9], a simple prediction of the power of the LR can be made, using the variance of true ICC and the approximate (ANOVA) sampling variance of their estimates from equation [4.14]. In this situation two different sources are expected to cause biases in the LR test; one source is that small samples cause departures of the distribution of the test statistic from the ChiSquare distribution, as in the previous section, the other source is that the estimates of the ICC are not normally distributed.

The power of a LR test to detect heterogeneity of ICC or phenotypic variances was investigated by simulation. Per replicate, the true t_i were sampled from a truncated normal distribution with mean t_0 (hence $t_i \sim N(t_0, v(t_i))$ in the interval <0,1>. For each of k herds, between and within sire SS were sampled from the appropriate X^2 distribution and the sample between and within sire components were estimated using REML. The sampling procedure caused a slightly skewed distribution of t_i since t_0 was 0.1. By sampling SS, data were assumed to be corrected for all fixed effects, including fixed herd effects.

For each of 5,000 replicates, LR tests were carried out corresponding to the following null hypotheses (H_0):

1) H_0 $[\sigma_0^2,\ t_0]$ = both ICC and phenotypic variances are homogeneous ($df = 2(k - 1)$);
2) H_0 $[\sigma_i{}^2,\ t_0]$ = ICC are homogeneous, allowing for heterogeneous phenotypic variances ($df = k - 1$);

3) H_0 $[\sigma_0^2, t_i]$ = Phenotypic variances are homogeneous, allowing for heterogeneous ICC ($df = k - 1$).

For each hypothesis the appropriate REML estimates were calculated using simple iterative techniques. The powers of tests 1 to 3 were predicted using:

$$P(\alpha) = \int_{[\ X^2_{\alpha}(df)\ ]/c}^{\infty} f(x)dx \quad , \quad f(x) \text{ being the density of a } X^2 \text{ distribution with } df \text{ degrees of freedom,} \qquad [4.15]$$

The constant $c = [\ (\ v(\hat{\theta}_i | \theta_i) + v(\theta_i)\ ) / v(\hat{\theta}_i | \theta_i)\ ]$ for hypotheses 2) and 3), with $\theta_i = t_i$ for hypothesis 2) and $\theta_i = \sigma_i$ for hypothesis 3). For hypothesis 1), $c = (c_2 + c_3)/2$, with c_2 and c_3 the constants for hypotheses 2) and 3) respectively.

4.2.3 Balanced cross-classified half-sib designs

4.2.3.1 Model specification

If sires and herds (strata) are cross-classified, i.e. all sires have progeny in all herds, then the following questions arise:

1) What is the contribution of the additional information, i.e. that animals in different herds are related to each other, to the detection of heterogeneity of parameters?

2) What is the effect of assuming a hierarchical design when maximising the likelihood, whilst data were generated from a cross-classified design?

The implicit assumption in the latter question, that data from individual herds were statistically independent of each other, was for example assumed by Swalve and Van Vleck (1987), Van Vleck and Dong (1988), Van Vleck *et al.* (1988) and in chapter 3, since relationships between animals in different herds were ignored in those studies. These questions were addressed again by using simulation. The following model was used to generate data consisting

of MSB (Mean Square Between sires within a stratum), MSW (Mean Square Within sires within a stratum) and MCPB (Mean Cross Product for sires between strata):

$$y_{ijl} = \alpha_i s_j + \beta_i e_{ijl} \quad , \qquad [4.16]$$

y_{ijl} is an observation on the l^{th} progeny (l=1,n) of sire j (j=1,s) in the i^{th} stratum (i=1,k) with residual e_{ijl}, and α_i and β_i are constants scaling the sire and residual variance. Therefore the assumption is that genetic correlations between sire performances in different strata are unity, and that a sire by herd interaction is the effect of scaling. Then, if **M** is a kxk matrix of MSB and MCPB between k strata and **W** is the diagonal matrix of MSW,

$$E[\, \mathbf{M}_{ii} \,] = \beta_i^2\sigma_w^2 + n\,\alpha_i^2\sigma_b^2 = \sigma_{wi}^2 + n\,\sigma_{bi}^2$$

$$E[\, \mathbf{M}_{im}] = n\,\alpha_i\alpha_m\sigma_b^2 = n\,\sigma_{bi}\sigma_{bm}$$

$$E[\, \mathbf{W}_i \,] = \beta_i^2\sigma_w^2 = \sigma_{wi}^2 \quad ,$$

for strata i and m. The likelihood function was parameterised in terms of between and within sire components, and was maximised conditional on the within sire within stratum mean square being the ML estimate (MLE) of the within component for that stratum, i.e. $MLE(\sigma_{wi}^2) = \mathbf{W}_i$. This was done for computational reasons (see next section). Although this parameterisation is different from the one used in the previous section, the main interest is the power of detecting heterogeneous between sire components, and this power is likely to be very similar to the power of detecting heterogeneous ICC. To verify this, a nested (hierarchical) design was simulated as in the previous section, but with parameterisation of the likelihood function in between and within sire components (see columns pertaining to O_4 in table 4.5). The effect of fixing the estimates of the within components to the within mean squares is unlikely to have a great effect on the likelihood ratio: even for the smallest design the degrees of freedom for the within components were as large as 270

(= 30*(10-1)).

4.2.3.2 Computing algorithm

Assume a matrix **M** of MSB and MCPB, and a diagonal matrix **W** of MSW, are observed from k herds (strata). Each of the s sires has n progeny in each herd (stratum). For the "full" model it is further assumed that:

$$E[\,\mathbf{M}\,] = \mathbf{V} = \mathbf{LL}' + \mathbf{D} \quad , \qquad [4.17]$$

Where **L** is a vector of length k with elements $\mathbf{L}_i = \sqrt{n}\,\sigma_{bi}$, and **D** is a diagonal matrix of order k with $\mathbf{D}_i = \sigma^2_{wi}$.

Then the residual likelihood is

$$-2L_u(\mathbf{M},\ \mathbf{W}|\mathbf{V}) = (s-1)[\,\log|\mathbf{V}| + \mathrm{tr}(\mathbf{M}\mathbf{V}^{-1})\,] + s(n-1)[\,\log|\mathbf{D}| + \mathrm{tr}(\mathbf{W}\mathbf{D}^{-1})\,] \qquad [4.18]$$

Conditional on $\mathbf{D} = \mathbf{W}$, and ignoring the second part of the likelihood pertaining to **D**, the maximum likelihood can be written as:

$$-2ML_u(\mathbf{M} \mid \mathbf{V},\ \mathbf{D}=\mathbf{W}) = (s-1)[\,\log(\theta_1) + \sum_{i=2}^{k} \theta_i + \sum_{i=1}^{k} \log(\mathbf{W}_i)\,] \qquad [4.19]$$

Where θ_i are the eigenvalues of $\mathbf{M}^* = \mathbf{D}^{-\frac{1}{2}}\mathbf{M}\,\mathbf{D}^{-\frac{1}{2}} = \mathbf{W}^{-\frac{1}{2}}\mathbf{M}\,\mathbf{W}^{-\frac{1}{2}}$ and θ_1 is the largest eigenvalue of $\mathbf{M}^*$.

Hence, conditional on $\mathbf{D}=\mathbf{W}$, no iterative procedure is required to calculate the maximum likelihood for the full model. Unless the number of strata is very large, calculating the eigenvalues for a symmetric kxk matrix is computationally relatively easy. The algorithm is similar to a commonly used algorithm in factor analysis, the analogy being to regard sires as the only "factor" in the analysis explaining the data (see e.g. Lawley and Maxwell, 1971).

Computation of the ML for the alternative hypothesis, that all sire

variances are the same, again assuming **D=W**, involves computing the ML estimate of σ^2_{b0}, the estimate of the overall sire variance. It can be shown that for the above model the ML estimate of σ^2_{b0} has an explicit solution, which is:

$$ML(\hat{\sigma}^2_{b0}) = [\ (\mathbf{1'D^{-1}MD^{-1}1}) - (\mathbf{1'D^{-1}1})\ ] \ / \ [\ n\ (\mathbf{1'D^{-1}1})^2\ ]$$

where **1'** is a row vector of length k with all elements unity.

If data from different strata are assumed independent, computations of the ML requires solving a cubic equation in σ^2_{b0}. The ML estimate of the common sire variance then satisfies, conditional on **D=W**,

$$\sum^{k} 1 \ / \ (\ \sigma^2_{wi} + n\sigma^2_{b0}\ ) = \sum^{k} \mathbf{M}_i \ / \ (\ \sigma^2_{wi} + n\sigma^2_{b0}\ )^2$$

Again, this is relatively straightforward to solve.

4.3 Results

4.3.1 Testing for equal group means with heterogeneous group variances

Table 4.1 shows the results using a LR test for small sample sizes using the fixed effects model. Clearly the LR is not distributed as a ChiSquare for small n, since the null hypothesis is more often rejected when it is true than was expected from the significance level. For example, for k = 10 and n = 10, the estimated probability of rejecting H_0 when H_0 was true, i.e. when the variance of true means ($v(\mu_i)$) was zero, was 10.8% at a nominal significance level of 5%. For n = 25, the predicted powers were close to the observed ones. Expanding the LR function typically gives a X^2 approximation exact to order $1/n$ (see e.g. Kendall and Stuart, 1973, pp. 234-272), so that the deviation of observed from predicted powers is not surprising for small n. Modification of the test statistic (e.g. Bartlett, 1937) would result in smaller differences between observed and predicted powers. In general, the observed powers were low. For $v(\mu_i)$ = 0.01, hence $CV(\mu_i)$ = 10%, the maximum power, 26.4%, was observed for k = 25 and n = 25.

Table 4.1: Statistical power (in %) for α=5% from a LR test for testing group means and allowing for heterogeneous group variances.

			k = 2		10		25	
n	$v(\mu_i)$	$v(\bar{x}_i)$	O	P	O	P	O	P
5	0	0.20	8.1	5.0	19.1	5.0	34.5	5.0
	0.01		8.8	5.6	21.5	6.5	39.4	7.3
	0.25		22.1	19.1	68.5	58.3	95.0	88.1
10	0	0.10	6.5	5.0	10.8	5.0	15.0	5.0
	0.01		7.9	6.2	14.9	8.1	23.1	10.2
	0.25		31.4	29.5	85.1	84.9	99.4	99.3
25	0	0.04	5.4	5.0	6.5	5.0	8.8	5.0
	0.01		8.8	8.0	16.5	14.0	26.4	21.5
	0.25		47.0	46.7	98.4	98.5	100	100

k,n= number of groups and observations within each group respectively.
O = observed power of LR test.
P = prediction of power LR test (from formula [4.9] in text).
Standard errors of observed powers were approximately 0.4%.

4.3.2 Balanced nested half-sib designs

Tables 4.2 and 4.3 show simulation results for small and medium group sizes for a balanced nested half-sib design. The coefficients of variation rather than the variances of the population parameters were displayed to make comparisons between the powers for t_i and σ_i^2. The design from table 4.2 was chosen to give similar standard errors of the heritability (= $h^2 = 4t$) estimates as were obtained in chapter 3 using field data. For the parameters used in table 4.2, the approximate standard error of the corresponding heritability estimate was 0.189 (from equation [4.14]). The probability of rejecting H_0 when it was true was very similar to the significance level for testing phenotypic variances and for testing heterogeneity of ICC. For the double homogeneity test the LR test detected heterogeneity even when one of the parameters, in this case the phenotypic variance, was homogeneous (see columns pertaining to O_1 in tables 4.2 and 4.3). Clearly the power for detecting heterogeneous ICC was very low compared with the power to detect differences in phenotypic variances. For example, if the $CV(t_i)$ in the population of herds was

0.3, which corresponds to a distribution of the heritability with mean 0.40 and standard deviation 0.12, then in approximately 37% of repeated samples of 25 herd estimates a difference in heritability would be detected. Table 4.3 confirms that some of the (small) differences between observed and predicted powers in table 4.2 were caused by small sample sizes. Again the difference in power between LR tests for t_i and σ_i^2 is striking. In general simulation results agreed well with their predictions.

Table 4.2: Observed (O_i) and predicted (P_i) powers (in %) for LR tests from a balanced half-sib design for k=25.

s=30, n=10, t_0=0.10, σ_i^2=1.0, α=0.05

CV(t_i)	CV(σ_i^2)	O_1	P_1	O_2	P_2	O_3	P_3
0	0	5.5	5.0	5.5	5.0	4.5	5.0
0.1	0	7.1	6.4	8.2	7.1	4.4	5.0
0.1	0.1	87.4	84.1	7.3	7.1	89.7	90.8
0.2	0	13.9	11.8	16.6	15.7	4.5	5.0
0.2	0.2	100	100	14.8	15.7	100	100
0.3	0	30.1	24.9	37.0	35.6	4.3	5.0
0.3	0.3	100	100	38.4	35.6	100	100
0.4	0	51.0	47.4	62.6	62.6	4.6	5.0
0.4	0.4	100	100	63.3	62.6	100	100
0.5	0	70.6	72.4	80.3	84.1	4.9	5.0
0.5	0.5	100	100	82.6	84.1	100	100

Range standard error (s.e.):
s.e.(O_1) 0 - 0.7%
s.e.(O_2) 0.4 - 1.2%
s.e.(O_3) 0 - 0.5%
Subscripts 1-3 refer to different null hypotheses:
1 = both ICCs and phenotypic variances homogeneous
2 = ICCs homogeneous, allowing for heterogeneous phenotypic variances
3 = phenotypic variances homogeneous, allowing for heterogeneous ICCs

Table 4.3: Observed and predicted powers for LR tests from a balanced half-sib design for k=10.

s=100, n=10, t_0=0.10, σ^2=1.0, α=0.05

$CV(t_i)$	$CV(\sigma_i^2)$	O_1	P_1	O_2	P_2	O_3	P_3
0	0	4.8	5.0	5.5	5.0	5.2	5.0
0.1	0	8.4	8.2	10.2	10.0	5.3	5.0
0.1	0.1	95.2	96.9	10.3	10.0	96.3	96.3
0.2	0	25.3	22.7	29.7	31.1	5.0	5.0
0.2	0.2	100	100	30.9	31.1	100	100
0.3	0	55.5	51.5	63.1	62.4	4.8	5.0
0.3	0.3	100	100	64.4	62.4	100	100
0.4	0	78.9	79.1	84.2	84.1	4.6	5.0
0.4	0.4	100	100	82.6	84.1	100	100
0.5	0	89.8	93.4	92.6	94.0	4.7	5.0
0.5	0.5	100	100	92.4	94.0	100	100

Range standard errors:
s.e.(O_1) 0 - 0.6%
s.e.(O_2) 0.3 - 0.8%
s.e.(O_3) 0 - 0.4%

In table 4.4 the predictions of the powers for large samples for two groups are shown. Such samples may be similar to estimating parameters from groups of herds which have been split according to the herd mean or herd variance. The standard error of the heritability is shown because results from studies investigating heterogeneity of variance in two or more groups (e.g. Hill *et al.*, 1983; Lofgren *et al.*, 1985; Dong and Mao, 1990) usually are reported in terms of differences between heritability estimates. Table 4.4 shows that even for large sample sizes moderate powers can be obtained using a LR test. For all sample sizes in table 4.4, the power of a LR test for detecting heterogeneity of phenotypic variances was 100%.

Table 4.4: Predicted powers (for α=5%) for detection of heterogeneous ICC in two groups for LR tests from various balanced half-sib designs, assuming phenotypic variances are homogeneous and t_0=0.10.

s	n	s.e.(h^2)	Power (in %)		
			CV(t)= 0.1	0.2	0.3
100	25	0.071	17	59	88
	50	0.061	22	72	94
	100	0.056	26	78	96
250	25	0.045	40	91	99
	50	0.039	52	96	100
	100	0.035	59	97	100
500	25	0.032	69	99	100
	50	0.027	80	99	100
	100	0.025	86	100	100
750	25	0.026	84	100	100
	50	0.022	91	100	100
	100	0.020	94	100	100
1000	25	0.022	91	100	100
	50	0.019	96	100	100
	100	0.018	97	100	100

4.3.3 Balanced cross-classified half-sib designs

Table 4.5 shows the results from simulating data from a balanced cross-classified design. Results are shown only for cases where $CV(\sigma_i^2) = 0$, i.e. $CV(t_i) = CV(\sigma_{bi}^2/\sigma_i^2) = CV(\sigma_{bi}^2)$. Hence between and within sire variances were heterogeneous, but their sum, the phenotypic variance, was the same for all herds. The first columns for each of the two population designs, i.e. columns O_4, can directly be compared with columns O_2 from tables 4.2 and 4.3. Clearly the power for detecting heterogeneous sire components and ICC are similar. The second column of observed powers in table 4.5 shows the effect of assuming the incorrect model for calculating the LR. The loss in power occurs because part of the information about the covariance structure of the MSB is not taken into account in the calculation of the Maximum Likelihood. Note that the estimates of the between and within components both for the unrestricted model (different between and within components for each stratum) and for

the H_0 hypothesis are unbiased (conditional on the ANOVA estimates for the between sire variance being positive), since the expectations of the mean squares in the usual ANOVA are not changed; ignoring the MCPB simply means that the variance of the estimates is increased. For a nominal significance level of 5%, the estimated type I errors for both designs were less than 1% if an incorrect model was assumed (columns O_5). The probability of rejecting H_0 when it was false, i.e. the power of the test, was also small when MCPB were ignored.

Table 4.5: Observed powers ± s.e. (in %) in detecting heterogeneous sire variances for LR tests from balanced nested and cross-classified half-sib designs.

	design I: k=25, s=30, n=10			design II: k=10, s=100, n=10		
$CV(\sigma_D^2)$	O_4	O_5	O_6	O_4	O_5	O_6
0	6.6 ± .3	0.4 ± .2	5.3 ± .3	5.7 ± .5	0.9 ± .1	5.1 ± .3
0.1	8.2 ± .3	0.5 ± .2	7.7 ± .5	10.0 ± .5	2.5 ± .3	14.1 ± .3
0.2	16.5 ± .6	2.1 ± .6	28.6 ± 1.1	29.4 ± .9	15.9 ± .4	49.9 ± .6
0.3	35.6 ± .9	12.2 ± .9	61.6 ± 1.5	58.5 ± .7	49.5 ± .5	82.8 ± .7
0.4	58.6 ± .4	32.2 ± 1.8	86.5 ± 1.1	80.8 ± .4	74.0 ± .7	94.7 ± .2
0.5	76.5 ± .6	57.8 ± 2.0	96.4 ± .6	91.3 ± .4	88.0 ± .7	98.5 ± .2

In all cases $CV(\sigma_i^2)$=0.
All LR are conditional on **D=W** (see text)
Subscripts 4-6 refer to the following data structures and hypotheses:
4 = data from nested design, H_0= homogeneous sire variances
5 = data from cross-classified design, but ignoring MCPB, H_0= homogeneous sire variances.
6 = data from cross-classified design, H_0= homogeneous sire variances.

The final column in table 4.5 indicates the gain of using MCPBs for the assumed model to detect heterogeneous variance components. The power was increased substantially, in particular for the range of $CV(t_i)$ of 0.2 to 0.3. In absolute terms, the power was still small for design I (25 strata, 30 sires, 10 progeny per sire): if the coefficient of variation of the between sire variance was 0.30 in the population, this heterogeneity would be picked up in approximately 62% of samples. For $CV(t_i)$ = 0.1, the power for the nested design (8.2%) was found to be larger than the power for the cross-classified design (7.7%) for the design with 25 herds, while a larger power was expected for the cross-classified design. This may be explained by

sampling (SE of mean powers were 0.3 and 0.5 respectively) and by departures from normality for small sample estimates. The estimated type I error for the nested design (column O_4) was 6.6%, at a nominal significance level of 5%, whereas the estimated type I error for the cross-classified design was 5.3%.

4.4 Discussion

The analytical and simulation results show clearly that the power of a LR test for detecting heterogeneous ICC (or heritabilities) is very low for the range of standard errors of h^2 estimates to be expected from individual herd data in most countries. In chapter 3, 6 years of first lactation data were used from 26 large pedigree herds in England and Wales, and standard errors of h^2 estimates of approximately 0.19 were obtained. Van Vleck and Dong (1988), using 300 to 400 first lactation records per herd, estimated the standard errors of their h^2 estimates to be approximately 0.15. The U.K. has the largest average herd size in Europe, so sampling variances of individual herd estimates would be larger in other countries in Europe. Using more records per herd seems obvious, but may give additional problems of heterogeneity of variance between herd-years and between lactations, if the use of later lactations was to be considered.

Therefore the conclusion from chapter 3, that h^2 estimates were fairly homogeneous and that phenotypic variances differed between herds, is not surprising given the low power of the statistical test. However, before using an AM-BLUP evaluation, a decision should be made with regards to the correct covariance structure of the data. Given the lack of power in detecting any differences in heritabilities between herds, it seems logical to assume that heritabilities are homogeneous. Records can then be scaled according to an (regressed) estimate of the within-herd phenotypic variances, if those variances were found to be heterogeneous. A Bayesian justification for assuming homogeneous h^2 is that the individual herd estimates should be regressed to an overall h^2 estimate (a prior for the mean of the distribution of the heritability) and since the sampling variances of the individual estimates are large, the

regressed estimates would be very similar (homogeneous). This regression is further investigated in chapter 7. Foulley *et al.* (1990) presented a general framework to test for sources (e.g. herds or sires) causing heterogeneity of residual variance, and presented an example to illustrate the generality of their test. However, the test failed to detect heterogeneity of residual variance caused by sires, and it may be argued that in the power of the presented hypothesis test, essentially a LR test, for detecting heterogeneity of sire variances (whether caused by herds or sires) is likely to be low in most practical situations. San Cristobal *et al.* (1990) questioned the robustness of their or any LR test to departures from normality, but the results from the first section, testing for equality of group means, and results from the half-sib designs suggest that for relatively small samples the lack of statistical power is of greater practical importance than violations of normality assumptions.

The power for large samples approaches unity rapidly (table 4.4), although differences in t (h^2) may not be detected for two herd-groups with 100-200 sires represented. For example, Hill *et al.* (1983) estimated parameters in two (high and low) groups, each with 762 sires and approximately 11 effective daughters per sire. Using the prediction formula [4.15], with $t = 0.0625$ ($h^2 = 0.25$) and $\alpha = 5\%$, repeated samples of 2 herd groups from the total population would give a power of 13, 32, 47, 58 and 65% for CV(h^2) = 0.1,0.2,...,0.5 respectively. These relatively low powers are confirmed by performing a simple t-test, now conditional on the estimates, on the difference of the estimates in the high and low group. Although the sign of the difference is consistent (high mean and high variance groups showed higher heritabilities), the test statistic is not significant at the 5% level.

Using information between herds or strata may increase the power of the LR test, but simplified models are necessary, for computational reasons, to make calculation of likelihoods under various hypotheses feasible. If, for example, in the cross-classified design the assumption about scaling was not made, the number of between sire parameters to be estimated would increase from k to $k(k+1)/2$.

To conclude, the power of detecting heterogeneous heritabilities or (additive) genetic variances between herds using field data is expected to be small, while it is relatively easy to detect differences in total phenotypic variances.

CHAPTER 5

UNIVARIATE AND MULTIVARIATE PARAMETER ESTIMATES FOR MILK PRODUCTION TRAITS IN LACTATIONS 1-3 USING AN ANIMAL MODEL. I: DESCRIPTION OF ANALYSES AND PRESENTATION OF REML ESTIMATES

5.1 Introduction

As discussed in section 1.2, assumptions about the covariance structure of observations analysed with a linear model are often simplified to make computations feasible. In particular, this is the case for prediction of breeding values for large populations, e.g. for a national evaluation. For example, the U.S.A (Wiggans *et al.*, 1988a), France (Ducrocq *et al.*, 1990) and Australia (Jones and Goddard, 1990) use a modified repeatability model for which a genetic correlation of unity is assumed between performances across lactations and some (pre)scaling is applied to later lactation records to account for higher phenotypic variances of traits in later lactations. Later lactation records are given lower weightings by adjusting the error structure of the observations, and milk, fat, and protein yield are analysed separately using this modified repeatability model. The potential loss in efficiency of selection by making these assumptions depends on the true, unknown, covariance structure of the data, and on the breeding goal. By estimating relevant (co)variances and assuming a particular combination of traits on which to select, the potential loss in efficiency of selection by using simplified covariance structures may be quantified.

For estimating (co)variance components it seems desirable to use the same model as is, or soon will be, used for the prediction of breeding values, i.e. an animal model. Few (co)variance estimates from AM analyses have been reported; Swalve and Van Vleck (1987) analysed milk yield in lactations 1-3, and Van Vleck and Dong (1988) performed a multivariate analysis on milk, fat and protein yield in the first lactation.

The aims of this study were:

1) To estimate multivariate (MV) parameters for milk (M), fat (F) and protein (P) yield in lactations 1, 2 and 3 (L1, L2, L3). Estimates of correlations between different traits in different lactations, for example between milk yield in lactation 1 (M1) and fat yield in lactation 2 (F2) have not been reported before. In the notation used, the number following M, F or P refers to lactation number, and the combination above, M1 and F2, may be written as M1F2. Similarly, a multivariate analysis on M1, F1 and P1 may be written as analysing M1F1P1.

2) To investigate the implications of the estimates for prediction of breeding values when simplified assumptions are made regarding covariances structures. This part of the study is presented in chapter 6.

5.2 Material

First, second and third lactation production records for the period 1979-1987 from 100 large pedigree herds were extracted from the Milk Marketing Board's production files. Herds were selected on the number of heifers present in 1987. Later lactation records, i.e. second or third, were included only from cows for which the previous lactations were present. All cows were pedigree Holstein-Friesian (HF). Some summary statistics of the data are presented in table 5.1. The data used to investigate heterogeneity of variance between herds (chapter 3) were a subset of the data used for this study.

Table 5.1: Summary statistics of data.

		LACTATION 1	2	3
Number of records		38811	26223	16542
Number of animal effects		58689	42835	28919
Number of sires		2357	1948	1565
Mean (kg)	M	5291	6143	6643
	F	208.8	239.7	257.8
	P	173.0	201.4	215.5
SD (kg)	M	1111	1335	1372
	F	44.6	53.0	55.7
	P	34.6	41.5	42.7

(SD = Standard deviation)

5.3 Methods

Residual Maximum Likelihood (REML; Patterson and Thompson, 1971) was used to estimate (co)variances, using programs based on software written by Meyer (1988, 1989). Fixed effects in the mixed linear model were herd-year-seasons (HYS) and month of calving. Seasons were defined as 4 month periods, corresponding to the definition used for the current U.K. sire evaluations. Proportion of Holstein-Friesian in the cow, age at calving and lactation length were fitted as covariables. All animal effects, including those of proven sires, were treated as random; this may cause a (downward) bias in the estimates, since comparisons between proven sires contribute to the estimate of genetic variance.

The following analyses were carried out:

1) Univariate analyses for each of M, F and P in lactations 1-3. If culling takes place on performance in previous lactations, the parameter estimates from univariate analyses on later lactations will be biased. Comparing variance components from these univariate analyses with components from models that (partly) take account of selection may give some indication about what kind of selection (if any) has acted on these data.

2) Analyses using a repeatability model for each of M, F and P in lactations 1 and 2. For this model it was assumed that the genetic correlation of performance between lactations was unity and that heritabilities were constant across lactations. A permanent environmental effect was fitted as an additional random effect for these analyses. Comparing results from these analyses with results from bivariate analyses may show how the (co)variances are partitioned when a genetic correlation of unity between performances in lactations 1 and 2 implicitly is assumed.

3) Within lactation (for L1, L2 and L3) MV analyses for traits M, F and P. An algorithm proposed Thompson and Hill (1990) was used to estimate (co)variances. Their algorithm was designed to reduce a multivariate estimation problem to a set of independent univariate estimations. Assuming equal design matrices for p traits, Thompson and Hill (1990) proposed performing $q=p(p+1)/2$ univariate analyses, where the q "traits" are obtained from linear transformations of the p traits, and suggested finding a transformation matrix (iteratively) that would stabilise the back-transformed pxp covariance matrix from one round to the next. Following Thompson and Hill's suggestion, the initial transformation matrix was chosen so that $p=3$ traits and $q-p=3$ sums of traits were analysed. Subsequently, after $q=6$ univariate analyses, a canonical transformation was calculated and 3 canonical variates were formed. The next "round" consisted of performing univariate analyses on these 3 canonical variates and on 3 pairwise sums of the canonical variates. The whole procedure was stopped after 5 complete rounds of iteration, since correlations on the original scale changed very little from round 4 to 5. Thompson and Hill (1990) proposed their algorithm for the general case of equal design matrices and more than two random effects in the linear model. For the analyses described above, only two random effects (animal and residual) were fitted, so that a "standard" canonical transformation (see e.g. Meyer, 1985) could have been applied. Both methods, however, should give similar estimates, since the described algorithm was found to be highly efficient (Thompson and Hill, 1990).

4) Bivariate (BV) analyses on all pairwise combinations of traits in different lactations. Unfortunately, analysing the data using a general MV model (for example with 3 traits in 3 lactations i.e. for 9 traits) was computationally not feasible. Therefore, selection bias is likely to affect some of the parameter estimates. In particular, (co)variances estimated for lactation 2 and 3 will be biased if culling was based on performance in the first lactation. For all BV analyses the fixed effect structure was different for both traits. Computations would be reduced if, for example, a particular fixed effect was assumed to be the same for M1 and M2, but this assumption is difficult to justify for other combinations (e.g. M1P3). For all BV analyses, the observations were scaled to their phenotypic standard deviation, since this was found to be more efficient when using a simplex algorithm (Meyer, 1989) to maximise the likelihood.

For most analyses data sets were too large to be handled in one single likelihood evaluation. Data sets were therefore randomly subdivided into subsets of herd groups. The estimates from each sample were assumed to be independent of other estimates. This assumption is strictly true, since some sires had progeny in different subsets. The correlation between estimates from different samples depends on the number of sires represented in different samples and their contribution to the parameter estimates in each sample. For analyses 1) and 2) data were split into 5 subsets of 20 herds each, for analyses 3) into 5, 4 and 2 herd groups (for L1, L2, and L3 respectively), and for 4) into 10 groups of 10 herds. For the univariate analyses and the analyses using a repeatability model, the standard errors (s.e.) of the estimates were calculated by approximating the likelihood surface at the maximum likelihood estimates by a quadratic function in the parameters of interest and using the matrix of second differentials to calculate asymptotic variances of the estimates (see chapter 3 for an application and discussion of this procedure). For the within lactation MV analyses and the BV analyses, the average (co)variance estimates are presented with the empirical standard error of the mean estimate. No weighting of estimates was applied because subsets were roughly of equal size

and there was insufficient information about the sampling (co)variances of the variance components (a weighting according to the number of records in the analysis was tried and showed differences between weighted and unweighted means of the order of 1% of the mean).

It was not clear how to combine the different estimates efficiently into one overall (9x9) covariance matrix, since there was insufficient information about sampling variances and culling bias. Estimates of variances and covariances of M, F and P in lactation 3, for example, were available from bivariate analyses L1L3 and L2L3 and from MV analyses within L3, all of which were probably subject to culling bias. The following method was chosen to create consistent 9x9 covariance matrices: For L1 the (co)variances from analyses 3) were used. The variances (diagonals) in L2 and L3 were taken from BV analyses L1L2 and L1L3 using the same trait in each lactation. For example, the variance estimate for P3 was used from analysis P1P3. Within lactation genetic and environmental covariances between M, F and P for lactations 2 and 3 were calculated using the variances as described above and the estimates of the within lactation genetic and environmental correlations. The phenotypic covariances were calculated as the sum of the genetic and environmental covariances thus created and phenotypic correlations were calculated from these. The same method was used to calculate covariances between different traits in different lactations, now using the genetic and environmental correlations estimated from BV analyses. This somewhat arbitrary way of combining different estimates was found to give fewest problems of negative definite covariance matrices. It was thus assumed that variances from BV analyses L1L2 and L1L3, and genetic and environmental correlations between traits within lactations, were least biased through selection.

To summarise the calculation of the 9x9 covariance matrices (presented in tables 5.5-5.7):

- All genetic, environmental and phenotypic (co)variances within lactation one were from multivariate analyses on M1F1P1.
- Environmental and genetic correlations between milk, fat and protein yield within lactations 2 and 3 were from multivariate

analyses on M2F2P2 and M3F3P3 respectively.

- Environmental, genetic and phenotypic variances for M2, F2, P2, M3, F3, and P3 were calculated from bivariate analyses on M1M2, F1F2, P1P2, M1M3, F1F3 and P1P3 respectively.
- Environmental and genetic correlations between traits between lactations were taken from bivariate analyses for each pairwise comparison.
- All remaining phenotypic covariances and phenotypic correlations followed directly from combining the above calculated elements.

Parameters for fat and protein content were approximated using a first order Taylor series expansion. If xi/yi and xj/yj are ratio traits in lactations i and j respectively, then an approximation of the covariance between those two traits is,

$$\mathrm{cov}(xi/yi, xj/yj) \approx$$

$$\frac{\mu_{xi}\,\mu_{xj}}{\mu_{yi}\,\mu_{yj}} \left[\mathrm{CV}(xi)\mathrm{CV}(xj)\, r_{xi,xj} - \mathrm{CV}(xi)\mathrm{CV}(yj)\, r_{xi,yj} \right.$$

$$\left. - \mathrm{CV}(yi)\mathrm{CV}(xj)\, r_{yi,xj} + \mathrm{CV}(yi)\mathrm{CV}(yj)\, r_{yi,yj} \right] , \qquad [5.1]$$

with CV the coefficient of variation ($= \sigma/\mu$) and $r_{x,y}$ the correlation between traits x and y. Formula [5.1] was applied using estimates of the coefficients of variation and estimates of the (co)variances for the yield traits in lactations 1-3.

5.4 Results

The main results of the different analyses are presented in tables 5.2-5.11. Heritabilities for production traits for the first lactation (table 5.2) were moderate to high. Although the genetic parameter estimates from the univariate analysis for lactation 2 may be biased through selection, the increase in the environmental

variance for lactation 2 (which is unlikely to be greatly affected by culling) was striking; the ratio of environmental variances in lactation 2 to that in 1 was approximately 1.6. Part of the increase in variances for the second lactation may be a scale effect (see also tables 5.1 and 5.8 for means and coefficients of variation), since the (biased) genetic variance for lactation 2 is also larger than the first lactation genetic variance.

Table 5.2: Univariate REML estimates for lactations 1-3 (variances in kg^2).

	M1	F1	P1	M2	F2	P2	M3	F3	P3
σ^2_a	238564	330.7	193.2	246425	349.9	218.9	207491	300.3	199.0
σ^2_e	371956	584.5	351.0	608266	950.2	571.3	693720	1121.6	647.1
σ^2_p	610520	915.2	544.2	854691	1300.1	790.2	901211	1421.9	846.1
h^2	0.39	0.36	0.36	0.29	0.27	0.28	0.23	0.21	0.24
s.e.(h^2)	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03

Results from analyses with a repeatability model are presented in table 5.3. Heritabilities were slightly lower than those estimated from univariate analyses on first lactations only. The variance component estimates from the analyses using a repeatability model (table 5.3) may be explained using the general bivariate model results from tables 5.5-5.7; it seems that both the genetic and phenotypic variances from the repeatability model were roughly the (weighted) average of the bivariate first and second lactation parameters, and the average environmental variance in lactation 1 and 2 was partitioned into a permanent environmental and residual variance. If selection were on first lactation performance, a repeatability model should account for this selection effect, conditional on a genetic correlation of unity between first and second lactation performance.

Table 5.3: Univariate REML estimates from first and second lactations using a repeatability model.

	M	F	P
σ^2_a	255855	355.1	214.5
σ^2_{pe} (permanent environment)	149167	224.9	152.6
σ^2_e	313138	509.6	295.8
σ^2_p	718160	1089.6	662.9
h^2	0.36	0.33	0.32
s.e. (h^2)	0.01	0.01	0.01
repeatability	0.56	0.53	0.55
s.e. (repeatability)	0.01	0.01	0.01

Table 5.4 shows the heritability and correlation estimates from the within lactation MV analyses. Heritabilities were similar to univariate (unitrait) estimates from table 5.2, as expected, and again heritability estimates from L2 and L3 are expected to be biased downwards. Phenotypic correlations between yield traits were very similar for different lactations, and genetic correlations were slightly lower in L2 in comparison with L1, but similar for L2 and L3. Genetic and phenotypic correlations between milk and protein yield were very high, and environmental correlations for these traits calculated using the estimates from table 5.4 were close to unity.

Table 5.4: Within lactation correlation matrices (x100) from within lactation MV analyses on M, F, and P.

	M1	F1	P1		M2	F2	P2		M3	F3	P3
M1	39	75	91	M2	28	62	88	M3	24	58	86
	(1)	(2)	(1)		(2)	(6)	(1)		(1)	(1)	(1)
F1	84	36	82	F2	82	25	75	F3	82	21	73
	(1)	(1)	(1)		(1)	(2)	(4)		(1)	(1)	(1)
P1	95	87	36	P2	94	87	26	P3	94	86	24
	(1)	(1)	(1)		(1)	(1)	(3)		(1)	(1)	(2)

For each 3x3 matrix: heritabilities (x100) on diagonals, genetic correlations above and phenotypic correlations below diagonals.
Empirical standard errors (x100) below each estimate.
Mean and empirical s.e. of parameter estimates were based on 5, 4 and 2 samples for L1, L2 and L3 respectively.

In tables 5.5-5.7 the combined 9x9 covariances matrices are

presented. The similarity between the various 3x3 lactation by lactation covariance blocks is striking. In a subsequent study the consequences of these results for prediction of breeding values are investigated further. From table 5.5 it seems that genetically L2 and L3 are essentially the same for the yield traits, with genetic correlations between performances in second and third lactations in excess of 0.97. Comparing pairs of covariances or correlations such as M1F2 and F1M2 shows that their values are similar, which indicates that the ratio of variances for traits in different lactations are similar for M, F, and P. Similar proportionalities seem to exist for environmental components (table 5.6). Environmental correlations between traits within lactations were similar for lactations 1-3. Phenotypic correlations between traits within lactations (table 5.7) are not necessarily the same as those from table 5.4, because of the way this table was constructed. Little change, however, is observed. Phenotypic correlations for M1M2, F1F2 and P1P2 were slightly higher than repeatability estimates from table 5.3. Again the proportionality of the various 3x3 covariance blocks is striking.

Table 5.5: Additive genetic covariance matrix (upper triangle) and genetic correlations (x100; below diagonals) for M, F and P in lactations 1-3.

	M1	F1	P1	M2	F2	P2	M3	F3	P3
M1	241594	6699	6276	227270	5893	6019	217083	5523	5808
F1	75	329.3	208.3	5931	316.8	198.4	5695	314.9	195.9
P1	91	82	196.0	5871	185.1	188.7	5736	182.1	189.4
M2	87	62	79	282462	6651	7289	272594	7174	7381
F2	59	86	65	62	410.1	236.0	7384	401.8	256.9
P2	78	70	86	88	75	244.0	7464	254.8	251.1
M3	84	60	78	98	70	91	274479	6164	7439
F3	55	85	64	66	97	80	58	415.8	243.3
P3	72	66	82	85	77	98	86	73	270.1

Parameters for L1 are from within first lactation MV analyses (consistent with first block from table 5.4).
Within lactation off-diagonals for L2 and L3 are calculated using variance components from BV analyses (diagonals) and correlations from within lactation MV analyses.
All other estimates are averages from BV analyses on 10 samples.
Range empirical s.e. (x100) of correlations: L1L2: 2 - 4
L1L3: 2 - 6
L2L3: 1 - 5

Table 5.6: Environmental covariance matrix (upper triangle) and environmental correlations (x100; below diagonals) for M, F and P in lactations 1-3.

	M1	F1	P1	M2	F2	P2	M3	F3	P3
M1	373134	13185	11098	193774	6446	6053	188310	5932	5379
F1	89	589.5	409.6	6589	298.8	228.0	6198	274.1	199.6
P1	97	90	353.7	5635	207.6	193.5	5300	186.9	168.4
M2	42	36	39	584290	20685	17481	263259	8532	7866
F2	35	41	37	90	914.2	648.6	8471	391.1	285.4
P2	42	40	44	97	91	559.3	7860	286.1	259.5
M3	38	32	35	43	35	41	644499	23051	18846
F3	30	35	31	35	40	38	89	1040.8	710.3
P3	36	34	37	42	39	45	97	91	591.8

Range s.e. (x100) of correlations: L1L2: 1 - 2
L1L3: 2 - 2
L2L3: 1 - 2

Table 5.7: Phenotypic covariance matrix (upper triangle) and phenotypic correlations (x100; below diagonals) for M, F and P in lactations 1-3.

	M1	F1	P1	M2	F2	P2	M3	F3	P3
M1	614728	19883	17373	421044	12339	12072	405393	11454	11187
F1	84	918.8	617.9	12520	615.6	426.4	11893	589.0	395.6
P1	95	87	549.7	11506	392.7	382.2	11035	369.0	357.7
M2	58	44	53	866752	27337	24770	535853	15707	15247
F2	43	56	46	81	1324.3	884.5	15853	792.9	542.3
P2	54	50	58	94	86	803.3	15324	540.9	510.5
M3	54	41	49	60	45	56	918978	29215	26286
F3	38	51	41	44	57	50	80	1456.6	953.6
P3	49	44	52	56	51	61	93	85	861.9

In table 5.8, heritability estimates for the 9 "traits" are given which are expected to be least biased through selection, with coefficients of variation for genetic, environmental and phenotypic effects. As before, lactations 2 and 3 seem very similar. For all yield traits the additive genetic CV slightly decreased from L1 to L2, and the environmental CV increased from L1 to L2. Scale effects therefore act differently for genetic and environmental effects, and there seems to be no single scale transformation which would standardise both genetic and residual variances across lactations.

Table 5.8: Heritabilities (x100), their empirical standard errors (x100) and coefficients of variation (CV; in %) from bivariate analyses.

	M1	F1	P1	M2	F2	P2	M3	F3	P3
h^2	40	37	36	33	31	30	30	29	31
s.e.(h^2)	2	1	1	2	2	2	2	2	2
CV_a	9.2	8.7	8.0	8.7	8.4	7.8	7.9	7.9	7.6
CV_e	11.5	11.6	10.8	12.4	12.6	11.7	12.1	12.5	11.3
CV_p	14.8	14.5	13.5	15.2	15.2	14.1	14.4	14.8	13.6

First and second lactation estimates are from M1M2, F1F2 and P1P2; third lactation estimates are from M1M3, F1F3 and P1P2. CVs are $\sigma_a/\bar{x}$, $\sigma_e/\bar{x}$ and $\sigma_p/\bar{x}$ respectively, using the means from table 5.1.

Many analyses that were carried out yielded different estimates for the same variance component. For example, an estimate for M1 was available from a univariate analysis, from a MV analysis with F1 and P1, and from 6 different BV analyses. All those different estimates for the same component are shown in table 5.9. For each row the two identical values were from within lactation MV analyses, since, for example, M1, F1 and P1 were analysed multivariately but pairwise combinations M1F1, M1P1 and F1P1 were not analysed bivariately. Diagonals in table 5.9 were from univariate analyses (see table 5.2). As expected, the various estimates for first lactation variances are very similar, since these estimates are free from selection bias. Ignoring first lactation information to estimate variances in later lactations reduces the additive genetic variances by approximately 10%, most likely due to culling bias. It is not clear why the highest estimate for any trait in L2 was from a combined analysis with the same trait in L1, i.e. M1M2 gave the highest estimate for M2, and F1F2 and P1P2 showed the highest estimates for F2 and P2 respectively. Using prediction equations for selection biases from Meyer and Thompson (1984), no selection strategy for first lactation production traits was found that would produce these results.

Table 5.9: Comparison of variance components estimates (kg^2) from different pairwise analyes.

Genetic components for trait 1, from combined analysis with trait 2

				Trait 2					
	M1	F1	P1	M2	F2	P2	M3	F3	P3
Trait 1									
M1	238564	241594	241594	242903	239838	240284	239325	235846	237094
F1	329.3	330.7	329.3	331.1	334.5	331.5	325.1	326.3	325.2
P1	196.0	196.0	193.2	195.1	194.7	196.0	194.4	192.5	192.8
M2	282462	266699	271265	246425	238205	238205	250612	241764	248120
F2	390.1	410.1	392.8	322.7	349.9	322.7	360.3	354.4	361.8
P2	242.3	238.5	244.0	207.6	207.6	218.9	221.7	215.3	224.4
M3	274479	244164	260351	255960	235948	246660	207491	216977	216977
F3	381.9	415.8	387.9	333.9	374.5	344.5	300.0	300.3	300.0
P3	262.2	248.3	270.1	241.6	235.7	249.2	204.2	204.2	199.0

Environmental components trait 1, from analysis with trait 2

				Trait 2					
	M1	F1	P1	M2	F2	P2	M3	F3	P3
Trait 1									
M1	371956	373134	373134	367511	369236	369113	368209	370676	369857
F1	589.4	584.5	589.4	582.7	579.4	582.8	585.7	584.8	585.0
P1	353.7	353.7	351.0	349.7	350.1	349.4	349.1	349.9	350.4
M2	584291	588395	588537	608266	603973	603973	587910	594193	589094
F2	916.2	914.2	917.3	961.3	950.2	961.3	919.6	925.0	918.4
P2	558.3	559.8	559.2	574.5	574.5	571.3	556.8	561.3	554.6
M3	644499	659592	652339	649053	658737	654627	693720	673182	673182
F3	1044.7	1040.7	1044.0	1074.8	1054.9	1072.1	1105.5	1121.6	1105.5
P3	594.7	605.4	591.8	603.7	606.7	601.4	633.4	633.4	647.1

A summary of the parameters calculated for fat and protein content (F% and P% respectively), from using equation [5.1], is presented in tables 5.10 and 5.11. Heritabilities for F% and P% were high and were fairly constant across lactations. Genetic correlations for F2%F3% and P2%P3% were substantially lower than the genetic correlations between yield traits in second and third lactations. Parameters for first lactation traits (M1, F1, P1, F1% and P1%) were similar to estimates from a 5x5 MV analysis on all traits in lactation one (results not presented). Genetic correlations between protein yield

and protein percentage were negative in first and positive in later lactations, although small in all cases.

Table 5.10: Parameters for fat content (F%) and protein content (P%) in lactations 1-3.

	F1%	P1%	F2%	P2%	F3%	P3%
F1%	58	60	85	54	84	52
P1%	37	62	45	78	44	76
F2%	55	34	64	64	74	48
P2%	33	56	40	63	49	78
F3%	52	34	62	38	61	62
P3%	35	57	39	60	42	62

Heritabilities (x100) on diagonals, genetic correlations (x100) above diagonals and environmental correlations (x100) below diagonals.

Table 5.11: Genetic and environmental correlations (x100) between yield and content traits within lactations 1-3.

	L1		L2		L3	
trait combination	r_g	r_e	r_g	r_e	r_g	r_e
F%M	-42	-22	-47	-20	-46	-17
F%F	28	25	40	26	46	30
F%P	-20	-13	-18	-10	-15	-6
P%M	-50	-36	-44	-34	-33	-37
P%F	-9	-18	11	-15	24	-17
P%P	-10	-10	4	-9	19	-12

5.5 Discussion

Univariate first lactation heritabilities were similar to the most recent U.K. estimates using a sire model (Meyer, 1987), but higher than estimates of Hill *et al.* (1983) and Meyer (1983 and 1984). Heritability estimates from pedigree populations are often higher than from non-pedigree populations (Meyer, 1987; Carabaño *et al.*, 1990). In dairy cattle, heritability estimates from daughter-dam regression are notoriously higher than estimates from paternal half-sib comparisons (Maijala and Hanna, 1974; Van Vleck 1986), and

since the AM-REML estimates are a combination of both, this may "explain" why the AM estimates are higher than previous estimates from sire models. Swalve and Van Vleck (1987) found AM-REML heritability estimates of approximately 0.33 for milk yield in the first three lactations, using a trivariate model and ignoring relationships between animals across herds. Information contributing to their heritability estimates were therefore mainly from daughter-dam comparisons. Van Vleck and Dong (1988) reported AM heritability estimates of 0.36, 0.35 and 0.33 for milk, fat and protein yield in first lactations. The increase of the phenotypic variance over time, additional to an increase associated with a higher mean production, is striking; a regression of the coefficients of variation (CVs) of milk production in the U.K. on time, using literature estimates from Hill *et al.* (1983), Meyer (1984 and 1987) and estimates from this chapter, shows a slight increase in the phenotypic CV from 1976-1987 and an increase in the genetic CV from 7% to 9%. The explanation for this observation is not clear, although perhaps better estimation procedures, in particular those accounting for selection on the data, may account for some increase in the estimate of the genetic variance in addition to a scale effect.

Genetic and phenotypic correlations between M1, F1 and P1 were slightly higher than the correlations found by Van Vleck and Dong (1988). Genetic correlations between M1, M2 and M3 were almost identical to the estimates of Swalve and Van Vleck (1987) and slightly lower than the sire model estimates of Meyer (1987). A small negative genetic correlation between protein yield and protein content in lactation 1 was also reported by Swanson and Gnanasakthy (1991). Genetic correlations between protein percentage and yield traits indicate that response to selection for fat and protein yield can be achieved without a reduction in the level of protein percentage, which accords with the wishes of many European dairy breeders. The explanation for the substantially lower genetic correlation between content traits in lactation 2 and 3, i.e. for M2%M3% and P2%P3%, compared with near unity correlations for the yield traits is not clear. Applying equation [1] to F2% and F3%, assuming all CVs are equal and genetic correlations for F2F3 and M2M3 are unity, gives,

$$r_{F2\%F3\%} = [1 - \tfrac{1}{2}(r_{F2M3} + r_{M2F3})] / \sqrt{[(1 - r_{M2F2})(1 - r_{M3F3})]}$$

Therefore one explanation may be that the within lactation correlations, calculated from within lactation MV analyses, were biased downwards relatively more than the between lactation between trait correlations which were calculated from BV analyses.

If culling of first lactation cows were on some linear combination of their milk, fat and protein production in the first lactation or on any "culling variate" correlated with the traits being analysed, this form of selection would only partially be accounted for when using a bivariate REML estimation (see Robertson (1966) for a detailed theoretical framework of a culling process). Therefore the BV second lactation parameter estimates may be slightly biased. The three traits considered were highly correlated, however, and the ratio of bivariate over univariate variance components was similar for all traits, which suggests that the bias may be small. Meyer and Thompson (1984) presented prediction equations of selection biases for a one-way sire classification, when culling was on a trait correlated with yield in the first lactation and maximum likelihood was used to estimate the parameters. Using their prediction formulas, the selection bias was investigated for various combinations of genetic and environmental correlations between the culling variate and the traits in the BV analyses. Selection intensity was calculated from the relative number of cows that had second lactations. It was found that for a range of parameter values likely to correspond with the true population values for milk, fat and protein yield, small biases were predicted for the estimates of the genetic parameters, but substantial biases (up to 40% of the true values) could occur for the environmental correlations between the two traits in the analyses. For example, if the culling variate was fat yield in lactation 1, the percentage biases in the estimate of the heritability for the trait in lactation 2 and for the genetic and environmental correlation would be 0, 0.4 and -4.4 respectively for M1M2, and 0.2, 0.3 and 4.4 for P1P2, using the BV parameter estimates as true population values. Although most of the information used in AM-REML is a combination of

comparisons between (paternal) half-sibs and daughter-dam pairs, the effect of selection on a correlated trait is unlikely to be large for the range of parameters investigated.

The parameter estimates from the bivariate model clearly showed that production traits in the second lactation are not repeated observations of first lactation records. Still, most countries use a repeatability model in their national AM evaluation, albeit with a lower weighting given to second and later lactation records. The weighting of later lactations seems the only instrument within the present day national AM evaluations to approximate the more appropriate multivariate model, for which heritabilities are lower and variances are much higher in later lactations. Additional to the implicit assumption of a genetic correlation of unity between first and later lactation yields, an improper weighting of later lactations when using a repeatability model will reduce genetic progress. Some calculations thereof are given in a subsequent study.

As described previously, the method used to create 9x9 covariance matrices from various available estimates was somewhat arbitrary. Any combination of estimates is expected to give sampling problems, since the traits are so highly correlated. For example, using heritability estimates from table 5.8 with genetic and phenotypic correlations from table 5.4 gives three within lactation environmental covariance matrices which all are negative definite. Using estimates of environmental correlations between M1, F1 and P1 from Maijala and Hanna (1974), Meyer (1985) and Van Vleck and Dong (1988), determinants of the environmental correlation matrix were found to be -0.003, 0.012 and 0.03 respectively, indicating that sampling problems may be expected with these traits. Still, when using the method described to calculate full 9x9 covariance matrices, sampling problems were not eliminated: the 9x9 genetic covariance matrix presented in table 5.5 is negative definite. However, the only negative eigenvalue is this matrix was relatively close to zero (-0.04 after standardising all phenotypic variances to 1.0 for M1, F1 and P1). Setting this eigenvalue to a small positive number (e.g. 10^{-6}) and recalculating all matrices showed very little difference for all variance components.

CHAPTER 6

UNIVARIATE AND MULTIVARIATE PARAMETER ESTIMATES FOR MILK PRODUCTION TRAITS IN LACTATIONS 1-3 USING AN ANIMAL MODEL. II: EFFICIENCY OF SELECTION WHEN USING SIMPLIFIED COVARIANCE STRUCTURES

6.1 Introduction

In chapter 5, genetic and environmental parameters were presented for milk yield (M), fat yield (F) and protein yield (P) in lactations 1-3. If the breeding goal for dairy cattle breeding is some (linear) combination of these production traits in all lactations, an optimal way to combine all available information to predict breeding values is a multivariate (MV) BLUP analysis. For a national animal model (AM) breeding value prediction, however, a general MV BLUP analysis is computationally not feasible. In practice, therefore, simplifying assumptions are made when predicting breeding values for large populations using an AM. In dairy cattle AM prediction, milk, fat and protein yield are usually evaluated separately using a repeatability model with some scaling for observations in later lactations to account for heterogeneity of variance across lactations (Wiggans *et al.*, 1988a and 1988b; Ducrocq *et al.*, 1990; Jones and Goddard, 1990).

In this chapter the loss in accuracy of selection is investigated when simplified covariance structures are used to predict breeding values, using selection index theory. A second aim is to investigate how to reduce the dimensionality of the above MV prediction problem to a managable size without a great loss in accuracy, using parameter estimates from chapter 5.

6.2 Material

As reported in section 5.5, the 9x9 genetic covariance matrix for M, F and P in lactations 1-3 was found to be negative definite. To create a (semi) positive definite matrix the single negative eigenvalue was set to 10^{-6}, and covariance matrices were recalculated. These matrices were then used for subsequent (index)

calculations. Without loss of generality, phenotypic variances for M, F, and P in lactation one were set to 1.0. The parameters are summarised in table 6.1.

Table 6.1: Scaled and rounded parameter estimates for milk, fat and protein yield in lactation 1-3.

	M1	F1	P1	M2	F2	P2	M3	F3	P3
M1	39	75	91	86	58	78	84	55	72
F1	89	36	82	61	86	70	60	84	66
P1	97	90	36	78	65	86	78	63	82
M2	42	36	39	33	62	88	95	64	83
F2	35	41	37	89	32	75	67	94	76
P2	42	40	44	97	91	31	88	77	96
M3	38	32	35	44	36	42	30	59	86
F3	30	35	31	36	41	38	89	29	73
P3	36	34	37	43	39	46	97	91	31
Vp	1.0	1.0	1.0	1.42	1.45	1.48	1.51	1.59	1.58

Heritabilities (x100) on diagonals, genetic correlations (x100) above and environmental correlations (x100) below diagonals.
Vp = phenotypic variance.

Table 6.2 shows the eigenvalues and eigenvectors of matrix $\mathbf{P}^{-1}\mathbf{G}$, where **P** and **G** are the 9x9 phenotypic and genetic covariance matrices of milk, fat and protein yield in lactations 1-3 (M1 F1 P1 M2 F2 P2 M3 F3 P3), calculated from parameters in table 6.1. As in chapter 5, a number following M, F or P indicates the lactation number. The smallest eigenvalue from the original $\mathbf{P}^{-1}\mathbf{G}$ was -0.03, and the corresponding eigenvector was

[-0.04 0.15 -0.17 0.40 0.36 0.19 -0.50 -0.50 0.09].

Hence the negative eigenvalue resulted mainly from the contrast of individual yield traits in lactations 1, 2 and 3 ((M2-M3) + (F1+F2-F3) + (P2-P1)). After setting the only negative eigenvalue of the original matrix **G** to "zero" (10^{-6}), the corresponding eigenvector for the newly formed matrix $\mathbf{P}^{-1}\mathbf{G}$ represented mainly the contrast between yield traits in lactations 2 and 3 (see last row of table 6.2). This was expected, given the very high genetic correlations for yield traits in lactations 2 and 3 (see table 6.1).

Table 6.2: Eigenvalues and eigenvectors of $\mathbf{P}^{-1}\mathbf{G}$ using estimates from table 6.1.

Eigenvalue	Corresponding Eigenvector								
	M1	F1	P1	M2	F2	P2	M3	F3	P3
0.68	0.87	-0.29	-0.48	1.41	-0.45	-0.91	0.30	-0.02	-0.29
0.63	-0.47	-0.17	0.72	2.29	-1.08	-1.20	-2.69	0.47	2.20
0.61	-0.72	-0.54	1.09	-1.62	0.10	1.47	1.42	-1.08	-0.29
0.51	4.35	-0.83	-3.29	-1.78	0.59	1.22	-1.42	-0.08	1.53
0.41	-1.65	0.71	1.38	1.17	0.83	-1.78	0.37	-0.81	0.70
0.36	-0.28	0.15	0.43	-1.10	-1.10	2.38	1.22	1.09	-2.21
0.15	0.20	-1.65	2.26	-0.07	0.78	-1.05	0.27	0.46	-0.52
0.11	0.35	2.11	-1.64	-0.27	-0.91	0.71	-0.13	-0.82	0.57
0.00	-0.06	0.01	-0.02	0.30	0.36	0.31	-0.37	-0.39	-0.16

6.3 Methods and results

For index calculations the following well known results were used (see, for example, Sales and Hill, 1976a and 1976b):

$$R = (\mathbf{a}'\mathbf{G}'\mathbf{P}^{-1}\mathbf{G}\mathbf{a})^{\frac{1}{2}} \qquad [6.1]$$

$$\hat{R} = (\mathbf{a}'\hat{\mathbf{G}}'\hat{\mathbf{P}}\hat{\mathbf{G}}\mathbf{a})^{\frac{1}{2}} \qquad [6.2]$$

$$R^{*} = (\mathbf{a}'\hat{\mathbf{G}}'\mathbf{P}^{-1}\hat{\mathbf{G}}\mathbf{a})(\mathbf{a}'\hat{\mathbf{G}}'\hat{\mathbf{P}}^{-1}\mathbf{P}\hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{a})^{-\frac{1}{2}} \qquad [6.3]$$

Where R, $\hat{R}$ and R^{*} are the optimal, predicted and achieved response to selection in the aggregate breeding value (= H) respectively, expressed as a ratio of the selection intensity. Further notation used,

$\mathbf{u}$ = qx1 vector of breeding values for q traits
$\mathbf{a}$ = qx1 vector of (marginal) economic values for q traits
$H = \mathbf{u}'\mathbf{a}$ = aggregate breeding value
$\mathbf{x}$ = px1 vector of sources of information on an individual (for

example phenotypic observations, daughter averages, predicted breeding values)

$\mathbf{b}$ = px1 vector of index weights

$I = \mathbf{b'x}$ = index value used to predict H

$\mathbf{P} = v(\mathbf{x})$; $\mathbf{G} = \text{cov}(\mathbf{x},\mathbf{u'})$ and the symbol $\hat{}$ added to a scalar of matrix indicates an estimate thereof.

Equations for the responses are from using

$\mathbf{b} = \mathbf{P}^{-1}\mathbf{Ga}$ for the optimal index, and

$\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{a}$ for an index using estimates of $\mathbf{P}$ and $\mathbf{G}$.

If a new trait is created which is a linear combination of the observations, $y = \mathbf{w'x}$, with $\mathbf{w}$ a px1 vector of weights, then the response to selection is

$$R = [\ \mathbf{a'G'w}\ (\mathbf{w'Pw})^{-1}\mathbf{w'Ga}\ ]^{\frac{1}{2}} \qquad [6.4],$$

and similarly, if $\mathbf{y} = \mathbf{W'x}$, i.e. variables in vector $\mathbf{y}$ are a linear combination of the variables in vector $\mathbf{x}$, then

$$R = [\ \mathbf{a'G'W}\ (\mathbf{W'PW})^{-1}\mathbf{W'Ga}\ ]^{\frac{1}{2}} \qquad [6.5],$$

It was assumed that the marginal economic value for any of the production traits in later lactations was the product of the relative expression of that trait and the phenotypic standard deviation, thus reflecting survival to later lactations and the economic importance of a larger standard deviation (and mean) in later lactations,

$\alpha_i = \epsilon_i\ \sigma_i$ where α_i , ϵ_i and σ_i are relative economic value, relative expression and standard deviation for lactation i.

Relative expression was assumed to follow a geometric series, $\epsilon_i = (0.8)^{i-1}$, assuming a relative survival of 80% from one lactation to the next and setting the expression in lactation one to 1.0. Phenotypic standard deviations were assumed to be 1.0, 1.20 and 1.25 for lactations 1-3, and 1.25 for all subsequent lactations. If it is further assumed that the covariance of any observation with the breeding value in lactation three is equal to the covariance of that

observation with breeding values in later lactations, i.e. the corresponding rows of matrix **G** are identical, then the economic values for lactations 1-3 are [1.0 1.0 4.0], since the sum of economic values for third and subsequent lactations is 4.0. Similarly for the case of considering only two lactations and assuming second and later lactation breeding values have equal covariances with observed phenotypes, **a** = [1.0 5.0]. Economic values for traits within a lactation were varied to reflect different breeding goals.

6.3.1 Single trait multiple lactations considerations

Meyer (1983) investigated the potential gain in response to selection from including multiple lactation information on progeny of sires for sire evaluation. The accuracy of selection was increased directly through more (genetic) information about the trait(s) of interest, and indirectly through a better data structure (better "connectedness"). Assuming **a'** = [1 1 4] for either milk, fat or protein yield in lactations 1-3, and using the relevant parameters for any of these traits from table 6.1, it was found (using standard selection theory) that for sires the increase in accuracy through including second (and third) lactation daughter information in the selection index was approximately 6%-10%. The number of progeny per sire for first and second lactations were varied from 25 to 50 and 5 to 35 respectively. See Meyer (1983) for more examples.

Perhaps a more interesting question regarding the use of multiple lactation information on a single trait is how much accuracy is lost when a repeatability model is assumed for breeding value prediction instead of the "true" MV covariance structure. This was investigated for three selection indices:

I1 = phenotypic index, i.e. sources of information are phenotypic observations on individuals

I2 = sire index; sources of information are daughter averages of sires in different lactations

I3 = cow index; sources of information are the predicted breeding value (index) of the cow's sire and dam and the cow's own records.

The largest reduction in response to selection is expected when

selection is across age classes, e.g. across cohorts with different amounts of information, since an improper weighting of later lactations then would have the largest impact. In the following examples, only 2 lactations and 2 cohorts were considerd, but the results are thought to be similar for more lactations (given the very high genetic correlation between second and later lactation yields) and more age groups. The genetic means for the cohorts were assumed to be zero, hence the consequences of the error in predicting genetic trend were ignored. A thorough study of long term losses in response through incorrect estimation of genetic trend (thus creating an suboptimal ranking of young vs. old animals) was outside the scope of this study. For each index there was different amounts of information on the two cohorts,

I1:
Cohort 1: phenotypic observation in lactation 1
Cohort 2: observations in lactations 1 and 2
I2:
Cohort 1: first lactation daughter average based on n_1 daughters
Cohort 2: n_1 first lactation daughter records and n_2 second lactation records ($n_2 < n_1$)
I3:
Cohort 1: sire index based on n_1 first lactation progeny, dam index based on sire index of dam and dam's first lactation record
Cohort 2: sire index based on $n_1 + n_2$ progeny records, dam index based on sire index of dam and dam's records in first and second lactation.

Parameters used for the example with 2 lactations and 2 cohorts were: $\mathbf{a}' = [1\ 5]$; $r_g = 0.85$; $r_p = 0.55$; phenotypic variances were 1.0 and 1.45 and heritabilities were 0.40 and 0.30 for first and second lactations respectively; $n_1 = 50$; $n_2 = 35$. The "estimated" (assumed) parameters were: $\hat{r}_g = 1.0$ (repeatability model); the true phenotypic covariance matrix was used and heritabilities for lactation 1 ($\hat{h}_1^2$) and for lactation 2 ($\hat{h}_2^2$) were varied. A proportion of 10% of the total number of animals available was selected. The definition of

repeatability model differs from the usual one because heritabilities and phenotypic variances are not necessarily equal in different lactations.

Responses to selection were calculated using equations [6.1], [6.2] and [6.3]. Given any set of parameters the optimal proportion of animals to be selected from each cohort was determined using an algorithm from Ducrocq and Quaas (1988), assuming the parameters used were the true population parameters. Results are presented in table 6.3. For the parameter set chosen the loss in efficiency was small; a 0%-5% reduction in genetic gain for a range of heritabilities for first and second lactation performance. These results may be expected, since the "true" genetic correlation (=0.85)between performance in lactations 1 and 2 was high and an observation for later lactation performance is always conditional on the presence of a first lactation observation. The ratio of achieved to predicted response was less robust to changes in parameters. Even when the correct heritabilities (0.40 and 0.30) were used the achieved response (accuracy) was approximately 10% below the maximum response. This may be seen as a very simple illustration that one should be cautious when using predicted breeding values (whether from selection indices or BLUP) to estimate genetic trend when the parameters used in the prediction are subject to large sampling errors or when they are a priori incorrect (as in the case of a repeatability model when it is known that $r_g<1$). Since results were similar for the three indices used, subsequent calculations were only performed for the case of mass selection.

Table 6.3: Relative responses to selection from using incorrect parameters in selection indices.

INDEX		I1			I2			I3		
$\hat{h}^2_1$	$\hat{h}^2_2$	α_1	R_1	R_2	α_1	R_1	R_2	α_1	R_1	R_2
20	20	42.9	100.0	153.2	48.7	99.0	121.3	46.6	97.9	131.1
	30	36.0	98.1	116.8	47.5	99.8	101.3	43.1	99.0	107.2
	40	30.0	94.9	91.9	46.5	100.0	88.4	39.7	98.1	90.2
40	20	49.0	97.8	106.3	49.9	97.1	104.9	49.5	98.7	107.2
	30	46.2	99.7	89.3	49.6	98.3	89.6	48.0	99.6	90.2
	40	42.9	100.0	76.6	49.3	99.0	79.7	46.0	99.3	78.2
60	20	50.0	95.0	81.5	50.0	95.8	96.6	50.0	95.8	89.1
	30	49.0	97.8	70.9	49.9	97.1	83.2	49.4	96.5	76.0
	40	47.3	99.3	62.9	49.8	98.0	74.6	48.3	96.4	66.8

$\hat{h}^2_1$ and $\hat{h}^2_2$ are heritabilities for first and second lactation performance used in selection index calculations.
$R_1 = 100(R^* / R)$ = achieved response as proportion of the maximum response.
$R_2 = 100(R^* / \hat{R})$ = achieved response as proportion of the predicted response.
α_1 = proportion of animals selected from age group 1.
Indices (see also text): I1 = index for mass selection
I2 = sire index
I3 = cow index
Economic values: **a** = [1.0 5.0]

6.3.2 Multiple trait multiple lactation considerations

Suppose the breeding goal is a linear combination of 9 traits (M1, F1, P1, M2, F2, P2, M3, F3, P3, where the number following M, F or P indicates the lactation number), which is thought to be a good indicator of lifetime economic production since second and later lactation performances are highly correlated. Then, choosing a set of economic values and using parameters from table 6.1, the relative accuracy of selection for different indices which use different amounts of information can be investigated. For three different sets of economic values these relative accuracies were calculated, and results are presented in table 6.4. The economic values for M1, F1 and P1 in the second breeding goal (H2) are similar to first

lactation economic weightings used in practical selection indices in Europe. H3 reflects a more "progressive" breeding goal with selection only on protein production. Results from table 6.4 show that approximately 20% accuracy is lost when only one observation is used to predict the aggregate breeding value. Results for H2 and H3 were similar since breeding values for these composite traits were highly correlated. If only accuracy is considered, using milk and fat yield in a selection index does not contribute substantially to increase response to selection for lifetime protein yield (H3).

Table 6.4: Accuracies of selection indices for mass selection as proportion (x100) of the accuracy using observations of M, F and P in lactations 1-3.

	BREEDING GOAL		
	H1	H2	H3
Traits fitted in selection index:			
M1	84.4	57.0	73.3
F1	79.1	71.2	63.9
P1	83.4	69.6	78.8
M2	80.2	57.6	74.1
F2	74.4	72.7	64.3
P2	81.8	74.1	82.0
M3	76.8	55.0	72.9
F3	70.1	71.8	60.2
P3	80.0	75.9	83.9
M1,F1,P1	85.9	78.4	79.4
M2,F2,P2	82.8	83.9	83.2
M3,F3,P3	80.3	89.2	87.8
M1,M2,M3	95.6	66.9	86.9
F1,F2,F3	89.8	86.1	75.3
P1,P2,P3	97.0	86.7	96.7
M1,F1,P1,M2,F2,P2	95.6	90.8	91.5

Breeding goals:
For H1, **a'** = [1 1 1 1 1 1 1 1 1]
H2, **a'** = [-1.1 1.0 2.2 -1.1 1.0 2.2 -4.4 4.0 8.8]
H3, **a'** = [0 0 1 0 0 1 0 0 4]

Although these calculations are an oversimplification of breeding value prediction and selection in practice, they are useful when comparing the accuracies from table 6.4 with accuracies when simplified assumptions are made regarding the covariance structure of the observations (in next section).

6.3.3 Proportionality considerations

One suggestion to reduce the dimensionality of a MV prediction problem is to investigate whether some traits may be approximately expressed as linear combinations of other traits, or if some linear combination of the traits explain most of the variation in the aggregate breeding value. To reduce computations (further) it would be of interest to find a minimum number of independent traits which would provide all the necessary information. In particular, it would be convenient if one linear transformation could be found that reduces the prediction problem of 9 highly correlated traits (milk, fat and protein yield in lactations 1-3) to that of 3 independent new traits.

Notation:

M = Moment matrix; a symmetric positive definite (PD) matrix of order lp, with mean squares and mean cross-products based on df degrees of freedom
l = number of lactations, p = number of traits per lactation
V = E(**M**); unknown PD covariance matrix of lp traits
K = symmetric matrix of proportionality constants of order l
$\otimes$ = direct product operator (see e.g. Searle, 1966), tr = trace operator
L = natural logarithm of likelihood.

Using standard multivariate theory (e.g. Anderson, 1958, chapter 10),

$$L(\mathbf{M},\mathbf{V}) = -\tfrac{1}{2}\, df\, [\, \log|\mathbf{V}| + \mathrm{tr}(\mathbf{M}\mathbf{V}^{-1})\,] \qquad [6.6]$$

$$= -\tfrac{1}{2}\, df\, [\, \Sigma \log\lambda_i + \Sigma \gamma_i\,] \quad \text{with } \lambda_i = \text{eigenvalue of } \mathbf{V}, \ \gamma_i = \text{eigenvalue of } \mathbf{M}\mathbf{V}^{-1}$$

The maximum likelihood (ML) is obtained for $\mathbf{V} = \mathbf{M}$,

$$ML(\mathbf{M},\mathbf{V}) = -\tfrac{1}{2}\, df\, [\, \log|\mathbf{M}| + lp\,] \qquad [6.7]$$

Suppose a moment matrix $\mathbf{M}_0$ is observed, and the null hypothesis is,

H_0: $\mathbf{V}_0 = E(\mathbf{M}_0) = k\mathbf{V}$ with $\mathbf{V}$ specified and k a constant.

Then,

$$L_0(\mathbf{M}_0,\mathbf{V}_0) = -\tfrac{1}{2}\, df\, [\ (lp)\log(k) + \log|\mathbf{V}| + (\Sigma\, \gamma_i)/k\ ] \qquad [6.8]$$

and the ML estimate of k, $\hat{k} = (\Sigma\, \gamma_i)/lp$. Hence

$$ML_0(\mathbf{M}_0,\mathbf{V}_0) = -\tfrac{1}{2}\, df\, [\ \log(\Sigma\, \gamma_i) + \log|\mathbf{V}| + lp\ ] \qquad [6.9]$$

For the trivial case of $\mathbf{M}_0 = k\mathbf{M}$ where $\mathbf{M}$ is the ML estimate of $\mathbf{V}$, all eigenvalues of $\mathbf{M}_0\mathbf{V}^{-1} = \mathbf{M}_0\mathbf{M}^{-1}$ are constant and equal to the proportionality constant (=k). The likelihood ratio (LR) test statistic, $t = 2(ML - ML_0)$, asymptotically has a X^2 distribution with degrees of freedom [$\frac{1}{2}lp(lp + 1) - 1$]. With observations on p traits in l lactations, one suggestion is to test $\mathbf{V}_0 = \mathbf{K}\otimes\mathbf{V}_h$, where $\mathbf{V}_h$ is a (transformation of a) submatrix describing a (co)variance block of p traits within or between lactations. For the case of M, F and P in lactations 1-3, the hypothesis is that the complete covariance matrix may be expressed as a proportionality matrix multiplied by a transformation matrix. $\mathbf{V}_0$ may be written as,

$$\mathbf{V}_0 = (\mathbf{K}_l \otimes \mathbf{I}_p)(\mathbf{I}_l \otimes \mathbf{V}_h) = (\mathbf{I}_l \otimes \mathbf{T}_p)(\mathbf{K}_l \otimes \mathbf{I}_p)(\mathbf{I}_l \otimes \mathbf{T}_p')$$

with $\mathbf{TT}' = \mathbf{V}_h$. Subscripts refer to the order of the matrices. Then,

$$L_0(\mathbf{V}_0,\mathbf{M}) = -\tfrac{1}{2}\, df\, [\ \log|\mathbf{K}\otimes\mathbf{V}_h| + \mathrm{tr}(\ \mathbf{M}(\mathbf{K}\otimes\mathbf{V}_h)^{-1}\ )\ ]$$

$$= -\tfrac{1}{2}\, df\, [\ l\log|\mathbf{K}| + p\log|\mathbf{V}_h| +$$

$$\mathrm{tr}(\ (\mathbf{K}^{-1} \otimes\mathbf{I}_p)(\mathbf{I}_l\otimes\mathbf{T}_p^{-1})\mathbf{M}(\mathbf{I}_l\otimes\mathbf{T}_p'^{-1})\ ) \qquad [6.10]$$

the trace in [6.10] may be written as

$$\mathrm{tr}(\ \mathbf{P}(\mathbf{K}^{-1} \otimes\mathbf{I}_p)\mathbf{P}'\mathbf{P}(\mathbf{I}_l\otimes\mathbf{T}_p^{-1})\mathbf{M}(\mathbf{I}_l\otimes\mathbf{T}_p'^{-1})\mathbf{P}')$$

$$= \mathrm{tr}(\ (\mathbf{I}_p\otimes\mathbf{K}^{-1})\mathbf{P}(\mathbf{I}_l\otimes\mathbf{T}^{-1})\mathbf{M}(\mathbf{I}_l\otimes\mathbf{T}_p'^{-1})\mathbf{P}')$$

$$= \mathrm{tr}((\mathbf{I}_l \otimes \mathbf{K}^{-1})\mathbf{M}^*)$$

$$= \Sigma \, \mathrm{tr}(\mathbf{K}^{-1} \mathbf{M}^*_{ii})$$

for $\mathbf{P}$ a permutation matrix and $\mathbf{M}^*_{ii}$ a lxl diagonal block of $\mathbf{M}^*$. Thus [6.10] becomes,

$$L_0(\mathbf{V}_0,\mathbf{M}) = -\tfrac{1}{2} df \, [\, l\log|\mathbf{K}| + p\log|\mathbf{V}_h| + \Sigma \, \mathrm{tr}(\mathbf{K}^{-1} \mathbf{M}^*_{ii}) \qquad [6.11]$$

Using [6.11], the ML estimate of $\mathbf{K}$,

$$\hat{\mathbf{K}} = (\Sigma \, \mathbf{M}^*_{ii})/l \qquad [6.12]$$

Unfortunately, the data from table 6.1 were found unsuitable for a LR test using equations [6.6] and [6.11]. Obviously the additive genetic covariance matrix (**A**) and the environmental covariance matrix (**E**) from table 6.1 are not independent moment matrices; **A** and **E** are highly correlated and the determinant of **A** is zero. One suggestion is to transform **A** and **E** into a between and within sire covariance matrix (**B** and **W**), assuming these matrices are from a balanced half-sib design based on s sires and n progeny per sire. However, there was insufficient information about the sampling variances of the estimated **E** and **A** matrices to determine the appropriate degrees of freedom. Furthermore, the exact distribution of the likelihood ratio test statistic based on empirically derived degrees of freedom and using animal model estimates may differ substantially from a Chi-Square distribution. Therefore, significance testing for proportionality was not persued.

Using parameter estimates from table 6.1, however, some inference with respect to proportionality may be drawn. One (obvious) choice for the transformation matrix in [6.10] is a canonical transformation on M1, F1, and P1. This transformation was calculated and the transformation matrix was used to transform the traits within second and third lactations.

Let $\mathbf{V}_{g11} = v_g \begin{bmatrix} M1 \\ F1 \\ P1 \end{bmatrix}$, and $\mathbf{V}_{p11} = v_p \begin{bmatrix} M1 \\ F1 \\ P1 \end{bmatrix}$

Then the transformation matrix of milk, fat and protein yield in lactation one, $\mathbf{Q}_1$, was chosen such that

$\mathbf{Q}_1 \mathbf{V}_{g11} \mathbf{Q}_1' = \mathbf{D}$ and $\mathbf{Q}_1 \mathbf{V}_{p11} \mathbf{Q}_1' = \mathbf{I}$

with elements $\mathbf{D}_i$ eigenvalues of matrix $(\mathbf{V}_{p11}^{-1}\mathbf{V}_{g11})$.
Using $\mathbf{Q}_1$, the vector of observations,
$\mathbf{y}'$ = [M1 F1 P1 M2 F2 P2 M3 F3 P3] = $[\mathbf{y}_1' \ \mathbf{y}_2' \ \mathbf{y}_3']$,
was transformed using:

$$\mathbf{y}_c = \mathbf{Q}\mathbf{y} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{Q}_1 \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix}$$

The eigenvectors for the 3 canonical variates in lactation 1 were [2.96 -0.72 -2.09], [-0.85 -1.85 2.61] and [-0.08 0.42 0.69] respectively, which form the rows of matrix $\mathbf{Q}_1$. The 9x9 correlation matrices and the heritabilities of the 9 new traits ($\mathbf{y}_c$) are shown in table 6.5. Off-diagonals in all 3x3 blocks were small, indicating that one transformation matrix created nearly three independent variates with for each transformed variate highly correlated observations in later lactations. Using the covariance matrix of milk, fat and protein yield in lactation 1 as $\mathbf{V}_h$, proportionality matrices for additive genetic and environmental effects were calculated from equation [6.12]. This assumed the observed covariance matrices **E** and **A** were moment matrices, but degrees of freedom needed not be specified. For **A** and E, the estimates of **K**, $\mathbf{K}_a$ and $\mathbf{K}_e$ respectively, were:

$$\hat{\mathbf{K}}_a = \begin{bmatrix} 1.00 & 1.03 & 1.06 \\ & 1.55 & 1.32 \\ \text{symm.} & & 1.74 \end{bmatrix}, \quad \hat{\mathbf{K}}_e = \begin{bmatrix} 1.00 & 0.60 & 0.58 \\ & 1.52 & 0.85 \\ \text{symm.} & & 1.69 \end{bmatrix}$$

Table 6.5: Parameters on transformed scale after applying the canonical transformation matrix from lactation 1 to lactations 2 and 3.

	C11	C12	C13	C21	C22	C23	C31	C32	C33
C11	64	0	0	79	5	-3	78	3	-8
C12	0	48	0	11	85	-3	10	82	1
C13	0	0	35	-5	8	85	-8	8	83
C21	64	2	-11	66	9	-15	77	4	-16
C22	4	38	11	0	52	8	4	71	10
C23	-3	0	43	-3	3	30	-13	11	98
C31	64	3	-1	71	6	-5	65	4	-24
C32	3	34	4	5	43	5	3	54	17
C33	-1	-2	36	-7	7	44	1	-6	29

Heritabilities (x100) on diagonals, genetic correlations (x100) above and environmental correlations (x100) below diagonals.
Cij = Transformed variate *j* in lactation *i*.

If proportionality is assumed, the 9x9 MV prediction problem may be reduced to three independent 3x3 multivariate predictions or to three independent evaluations with a repeatability model. Using the breeding goals defined previously the efficiency of this reduction in dimensionality was calculated for mass selection, conditional on the parameters in table 6.1 being the true population parameters. Thus the parameters from table 6.5 were used with all off-diagonals of all 3x3 covariance blocks set to zero. In the case of a repeatability model on the canonical variates, genetic correlations between canonical variates across lactations were set to unity. Results of selection index calculations for phenotypic selection are presented in table 6.6. The relative accuracy when using the first three canonical variates is slightly lower than the corresponding accuracy using the original first three variates (M1, F1 and P1) from table 6.3 because the genetic covariance structure between the canonical traits in lactation one and transformed variates in later lactations was simplified (off-diagonals of 3x3 blocks in matrix **G** were set to zero). Clearly little accuracy is lost assuming proportionality of the covariance structure for milk, fat and protein yield across lactations. Simplification to a repeatability model on three canonical variates was approximately 97% as efficient compared to a multivariate analysis on 9 traits. When using the canonical variates there was no advantage of a MV analysis over an analysis with a

repeatability model.

Table 6.6: Accuracies of selection indices for mass selection as proportion (x100) of the accuracy using observations of M, F and P in lactations 1-3, assuming proportionality of covariance blocks between traits across lactations.

	BREEDING GOAL		
	H1	H2	H3
Multivariate model using traits:			
C11,C12,C13	85.8	78.2	79.1
C21,C22,C23	82.0	83.0	82.3
C31,C32,C33	78.5	87.1	85.4
All transformed variates	98.5	97.8	97.6
Repeatability model using traits:			
C11,C12,C13	85.7	78.3	79.1
C21,C22,C23	81.6	83.5	82.2
C31,C32,C33	78.1	87.2	85.5
All transformed variates	98.2	96.4	96.7

Breeding goals:
For H1, **a'** = [1 1 1 1 1 1 1 1 1]
H2, **a'** = [-1.1 1.0 2.2 -1.1 1.0 2.2 -4.4 4.0 8.8]
H3, **a'** = [0 0 1 0 0 1 0 0 4]
Cij = Transformed variate *j* in lactation *i*.

6.3.4 Analysing linear combinations of the observations

A final reduction in dimensionality is achieved by analysing a reduced set of traits which are linear combinations of the available observations. One suggestion is to create a single new trait which is the sum of the phenotypic observations weighted by the corresponding economic values in the aggregate breeding value. Using the notation from equation [6.4], $y = \mathbf{a'x}$, where variables in **x** are, for example, observations for M1, F1 and P1. Relative accuracies were calculated using equation [6.4], fitting first lactation yield traits, first and second lactation yield traits, and all yield traits in vector **x**, respectively. Results are presented in table 6.7.

Table 6.7: Accuracies of selection indices for mass selection as proportion (x100) of the accuracy using observations of M, F and P in lactations 1-3, when using linears combinations of the observations as traits.

	BREEDING GOAL		
	H1	H2	H3
$\underline{y = w'x}$			
$w = a$, using traits			
M1,F1,P1	85.7	76.5	78.8
M1,F1,P1,M2,F2,P2	95.0	88.9	90.7
M1,F1,P1,M2,F2,P2,M3,F3,P3	98.1	92.0	92.8
$\underline{y = W'x}$			
$y_i = a_i'x_i$, using traits			
y_1, y_2	95.3	88.9	90.7
y_1, y_3	95.2	92.2	93.5
y_2, y_3	90.9	91.7	92.4
y_1, y_2, y_3	98.9	95.5	96.7

Breeding goals as in tables 6.4 and 6.6
$w = a$: weights are economic values
x_i = [Mi Fi Pi] (for i=1,2,3): milk, fat and protein yield for lactation *i*

Another suggestion is to use linear combinations of the yield traits within a lactation as new traits and to perform an analysis on those new traits. For example, if $y_1 = w_1'x_1$, for x_1'=[M1 F1 P1], and $y_2 = w_2'x_2$, for x_2'=[M2 F2 P2], then in the selection index framework this would be fitting $y = W'x$ as used for equation [6.5]. Using x_1 and x_2 as above, and x_3' = [M3 F3 P3], 3 new traits were created using the economic values for each trait in the aggregate breeding value as elements for matrix **W**. Accuracies for fitting combinations of these new traits are presented in table 6.7.

Comparing results from tables 6.4, 6.6 and 6.7 shows that little efficiency was lost when analysing linear combinations of the observations using economic values as weights. For the case of just using observations for M1, F1 and P1 this is not surprising, since these traits were so highly correlated and had similar heritabilities, hence their index values resembled the economic values. For breeding goals H2 and H3 approximately 8% accuracy was

lost when using all observations weighted by their economic values as a single trait, and approximately 4% accuracy was lost when analysing 3 new traits, each trait being a linear combination of observations and economic values within a lactation (table 6.7).

6.4 Discussion

Only one aspect of efficiency of selection, namely accuracy of predicting some aggregate breeding value assuming fixed effects were known, was considered in this study. Meyer (1983) found that for BLUP prediction of breeding values increase in accuracy from including later lactation observations was largely through an improved data structure. All results should therefore be seen as a first order approximation. Results from including information from relatives in the calculations, and including comparisons between young and old animals should have more direct relevance to practical breeding programmes.

Assuming a repeatability model for milk production traits across lactations seemed to have little effect on accuracy of selection, although the predicted gain/accuracy may be approximately 10% too high. More research is needed to investigate long term losses in response to selection when incorrect models are used to predict breeding values.

More information on the sampling variance of the parameter estimates are needed for testing the proportionality hypothesis. Ideally, one MV REML analysis on the 9 traits should be carried out, with an algorithm that would produce (2^{nd}) derivatives. Still, calculations then would involve a 90x90 (45 genetic and 45 environmental) sampling variance matrix which would probably be subject to large sampling errors itself.

As pointed out by Meyer (1985), the canonical variates from creating independent variates in lactation 1 may have a biological explanation. The eigenvectors show that canonical variate 1 corresponds approximately to percentage protein (and fat content to a lesser extent) and canonical variate 2 to the difference between fat

and protein content. Canonical variate 3 seems just to be the sum of fat and protein yield. Heritabilities for canonical variates were consistent with heritabilities found for fat and protein content previously reported. Canonical variates from diagonalising the complete (9x9) $\mathbf{P}^{-1}\mathbf{G}$ matrix have similar biological explanations as the canonical variates from lactation 1, but now including comparisons between lactations (see table 6.2).

Given that parameter estimates from table 6.1 are subject to sampling error, matrices describing covariances between M, F, and P within and between lactations were remarkably proportional to each other. Calculations for mass selection confirmed that little information is lost if proportionality is assumed. A repeatability model on canonical variates from lactation one should account for selection bias and only loses approximately 3% in accuracy compared to a general multivariate prediction of breeding values of milk, fat and protein yield in lactations 1-3.

Reducing the dimensionality of the prediction problem by analysing linear combinations of observations and economic values of corresponding breeding values, was found to be very efficient. However, no information from relatives was included in the calculations, and for the traits considered heritabilities and phenotypic and genetic correlations were similar between pairs of traits. When using traits with genetic and environmental correlations with opposite signs, and including observations over time, then if BLUP is used to calculate breeding values this method of creating new traits from the available observations is expected to be less efficient.

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSIONS

In chapter 1 genetic and statistical models were discussed which underlie most animal breeding data analyses. In general a balance has to be struck between the most desirable ("true") model and a model which is practical to implement. Finite (computer) resources and insufficient knowledge about parameters needed for more realistic biological models, e.g. lack of information about the number and action of genes that influence quantitative traits, usually result in many simplifying assumptions being made. The genetic model (implicitly) assumed in practice for prediction of breeding values and estimation of variance components is the infinitesimal model (see section 1.2.3), for which breeding values follow a normal distribution. The statistical model is usually a linear model with (multivariate) normality assumed for data and all random effects, since unbiasedness properties of BLUP hold under these assumptions for selected populations (section 1.2).

In dairy cattle the use of Best Linear Unbiased Prediction and normal distributions of random effects is hardly questioned (except, for example, Dempfle and Grundl, 1988; Gianola *et al.*, 1988; Gianola, 1990); but a national BLUP evaluation requires additional assumptions to make computations feasible, and justification for these extra assumptions is not always given. For example, the covariance structure for traits across lactations, i.e. a genetic correlation less than one and lower heritabilities and higher phenotypic variances in later lactations, is usually simplified for computational reasons. The tendency is to precorrect records for heterogeneous phenotypic variances across lactations and to assume a repeatability model combined with a weighted analysis, for which later lactations have weights of approximately 0.8 (Bonaiti and Boichard, 1990; Jones and Goddard, 1990). The exact model that is fitted is, however, not clear, since different error variances for later lactations seem not to be taken into account when the (mixed model) equations are set up. Therefore, it is not known how efficient such national evaluations are in predicting breeding values

for first lactation or lifetime yields. Gianola (1986), Weller (1988) and Quaas *et al.* (1989) presented models to take account of heterogeneity of variance, for example across lactations, for practical BLUP analyses, if heterogeneity of variance is assumed to be a scale effect. Further research is needed to investigate the effect of simplified models, and to find an optimum balance between (expected) computer power and the efficiency of predicting breeding values. A logical criterion may be to compare (computer) costs with benefits (genetic progress) for alternative models. Extending the selection index approach used in chapter 6, by including information from relatives and comparing animals over time, as for example in Wray and Hill (1989), may give some answers about the efficiency of simplified models. Future research, using, for example, simulation, also may show how robust the genetic (infinitesimal) model is for predicting long term responses to selection if more realistic assumptions are made with respect to (changes in) gene frequencies, gene numbers and gene actions affecting quantitative characters.

In chapter 2 it was shown that if variances are estimated within herd-mean groups, this is unlikely to give biased variance estimates caused by correlations between sire progeny means and herd means. Perhaps a more interesting question is how to deal with subpopulations within the (conceptual) whole population. Subpopulations are, for example, pedigree vs. non-pedigree herds, high vs. low yielding herds, intensively farmed vs. extensively farmed herds, and nucleus herds. Different parameters, such as heritabilities and phenotypic variances, are frequently found in different subpopulations. For the U.K. Holstein-Friesian population, Meyer (1987) estimated parameters for milk production traits in pedigree and non-pedigree herds and Hill *et al.* (1983) contrasted parameter estimates for milk production traits in high vs. low yielding herds. Brotherstone and Hill (1991a and b) compared survival in several subpopulations and found substantial differences for genetic parameters in different subpopulations. A relevant question is in which subpopulation genetic progress should be made and how this may be achieved, in particular if genetic correlations between performances in different subpopulations are not one. Fortunately, estimates of genetic correlations for milk production traits between

subpopulations are often close to unity (Hill *et al.*, 1983; Carabaño *et al.*, 1990; Dong and Mao, 1990; Short *et al.*, 1990). Hence, if subpopulations can be identified easily, a simple scaling of observations may be sufficient to correct for heterogeneity of variance between subgroups. If heritabilities differ substantially in different subpopulations, this should be taken into account in prediction of breeding values (Schaeffer *et al.*, 1978; Gianola, 1986; Quaas *et al.*, 1989).

In chapter 3 it was concluded from a sample of data from 26 pedigree herds, that heritabilities were homogeneous and phenotypic variances heterogeneous between individual herds. It was shown in chapter 4 that the statistical power of a likelihood ratio test as was used in chapter 3 is very low. Hill (1984) and Brotherstone and Hill (1986) proposed to regress individual herd parameters to an overall (prior) estimate, a standard Bayesian procedure in which the regression coefficient depends on the sample variances of individual herd estimates and the variance of the parameters. In general, their regression may be written as:

$$\hat{\theta}^{*}_{i} = \hat{\theta}_{0} + \beta_{i}(\hat{\theta}_{i} - \hat{\theta}_{0}) \qquad [7.1]$$

Where $\hat{\theta}_{i}$ and $\hat{\theta}_{0}$ are the parameter estimate for herd *i* and the overall (prior) estimate respectively. $\hat{\theta}^{*}_{i}$ is the regressed parameter estimate for herd *i*.
β_{i} is the regression coefficient, $\beta_{i} = 1 / (1 + \lambda_{i})$,
with $\lambda_{i} = var(\hat{\theta}_{i} | \theta_{i})/var(\theta_{i})$, the ratio of the sampling variance and the variance of the parameter, or less formally, the ratio of variance "within" and between parameters θ. Brotherstone and Hill (1986) suggested estimating $var(\theta_{i})$ by:

$$\text{variance between } \theta_{i} = \text{empirical variance between } \hat{\theta}_{i} - \text{average sampling variance of } \hat{\theta}_{i} \qquad [7.2]$$

An alternative way to estimate $v(\theta_{i})$ is to use the likelihood ratio (LR) statistic from the comparison between the maximum likelihood from estimating a single heritability estimate using all data (=

ML_0), and the sum of the maximum likelihoods from estimating individual herd heritabilities (= ΣML_i). If L_i is the likelihood function for parameter θ from herd i, and assuming this function is quadratic in θ, then

$L_i = a_i + c_i\ (\theta_i - \hat{\theta}_i)^2$, with a_i and c_i constants and $\hat{\theta}_i$ the maximum likelihood estimate of θ_i , and

$$2(\Sigma ML_i - ML_0) = - \Sigma\ 2c_i(\hat{\theta}_i - \hat{\theta}_0)^2 \text{ , with } \hat{\theta}_0 = ML(\theta_0)$$

Taking $\hat{\theta}_i = \theta_i + \epsilon_i$; $\hat{\theta}_0 = \theta_0 + \epsilon_0$; $v(\epsilon_i) = v(\hat{\theta}_i | \theta_i)$,

the expectation of (twice) the difference between the two maximum likelihoods is, approximately,

$$\begin{aligned} & E\ 2[\ \Sigma ML_i - ML_0\ ] \\ & = E\{\ \Sigma\ [\ (\epsilon_i - \epsilon_0)^2 / v(\epsilon_i)\ ] + \Sigma\ [\ (\theta_i - \theta_0)^2 / v(\epsilon_i)\ ]\ \} \\ & \approx \quad df + v(\theta_i)[\ \Sigma\ 1/var(\epsilon_i)\ ] \end{aligned} \qquad [7.3]$$

An estimate of $v(\theta_i)$ is, therefore,

$$\hat{v}(\theta_i) = (t - df)\ /\ [\ \Sigma\ 1/var(\epsilon_i)\ ] \qquad [7.4]$$

with $t = 2(\ \Sigma ML_i - ML_0\ )$, and df = degrees of freedom.

Using the heritability estimates for fat yield and their standard errors from table 3.2, the empirical variance between the 26 heritability estimates and the average sampling variance were found to be 0.035 and 0.039 respectively. Hence their difference, using [7.2], was negative. These values confirm the outcome of a more elaborate likelihood ratio test, i.e. that heritabilities were not significantly different from each other. Suppose there were true differences between heritabilities, but that sampling variances were

relatively too large to detect them (see also chapter 4 for statistical power calculations). What would the effect be of the regression as proposed by Brotherstone and Hill (1986)? One suggestion is to assume a coefficient of variation (CV) for heritabilities between individual herds, and to apply equation [7.1]. Using the mean estimate of the 26 estimates as an estimate of the population value (the "prior" estimate), i.e. as an estimate of the mean of all true individual herd heritabilities, and assuming a CV of 5%-20%, equation [7.1] was applied to the 26 heritability estimates of table 3.2. Results are presented in table 7.1. For a CV of 10% the average regression was 0.05, and the standard deviation of the regressed heritabilities was 0.01. Hence, if the CV is low, as suggested by the likelihood ratio test and the approximate estimate using [7.2], it seems debatable whether it is worthwhile to estimate individual herd heritability estimates for many herds, since the regressed values are nearly homogeneous.

Table 7.1: Regressions of individual herd heritability estimates to an overall mean, assuming different coefficients of variation for the unobserved heritabilities.

CV	$\bar{\lambda}$	$\bar{\beta}$	$\bar{\theta}^*$	sd(θ^*)	range(θ^*)
0.05	102.5	0.013	0.387	0.003	0.379 - 0.392
0.10	25.6	0.049	0.386	0.011	0.356 - 0.406
0.15	11.4	0.103	0.384	0.023	0.325 - 0.425
0.20	6.4	0.167	0.382	0.037	0.292 - 0.449
0.25	4.0	0.239	0.381	0.052	0.260 - 0.476

CV = coefficient of variation (σ/μ)
λ and β are variance ratio and regression, from [7.1]
θ^* = heritability after regression
sd(θ^*) = standard deviation of regressed heritabilities
$v(\theta) = (CV)^2(\hat{\theta}_0)^2$, with $\hat{\theta}_0 = \bar{\theta} = 0.388$

Using the likelihood ratio statistic from table 3.2, the standard deviation of heritabilities was estimated using equation [7.4], and was found to be 0.098, corresponding to a CV of 25.3%. The relatively large difference between the estimates of the variances using [7.2] and [7.4] may be explained by different weightings used in those formulas: [7.2] gives equal weightings to all individual herd

heritability estimates, whereas in [7.4] the heritability estimates are weighted according to their sampling variances. The last row of table 7.1 shows the parameters from using regression [7.1], assuming a CV of 25.3%. Even for the estimate of the CV of 25%, the standard deviation of regressed heritabilities was only 0.05, and data were from large pedigree herds. Smaller herds would give heritability estimates with such large standard errors that the regressed values would be very close to the overall mean (or prior estimate) and heritabilities could be regarded as being homogeneous. Some loss in accuracy of selection occurs if homogeneity of heritabilities is assumed when in fact true differences exist. Hill *et al.* (1983) showed, using selection index calculations, that for selecting sires across groups (herds) the optimal weight for a progeny mean from herd i is proportional to $n_i\sigma_{bi}/\sigma^2_{wi}$, where n_i, σ_{bi}, and σ^2_{wi} are the number of progeny, the between and within sire variance in herd i respectively. Assuming a constant phenotypic variance across herds (i.e. after scaling of observations), expectations of accuracies of sire selection were calculated using either the correct weights from above, or using $n_i\sigma_{b0}/\sigma^2_{w0}$ as weights, with σ_{b0} and σ^2_{w0} the average variances over all herds. The difference between the expectations of accuracies was found to be negligible. Similar results are obtained for mass selection across groups, using results from Hill (1984).

The two methods to estimate $v(\theta_i)$ were applied to the phenotypic variances, and were found to give similar results. Using equations [7.1] and [7.2], a CV of 24.8% was estimated (from a mean of 835 kg^2) and the standard deviation of regressed variances was 189 kg^2. The average regression coefficient was 0.86. Using equation [7.4], a CV of 22.8% was estimated, resulting in $sd(\theta^*)$ of 190 kg^2 and an average β of 0.83. The sampling variances of estimates of the phenotypic variances were calculated from the sampling (co)variances of genetic and environmental variances. Using further approximations, equation [7.4] was applied to the phenotypic standard deviation. The mean and CV of phenotypic standard deviations were estimated and were found to be 28.6 kg and 12.4% respectively. After regressing individual herd phenotypic standard deviations (average β was 0.86), the sd of the regressed values was 3.4 kg. Comparing the estimates of the CVs for phenotypic variances and heritabilities again shows that the findings

from chapter 3, homogeneous heritabilities and heterogeneous phenotypic variances, merely follows from a lack of statistical power, and does not necessarily say anything about true differences between heritabilities and between phenotypic variances. For practical purposes, it is concluded that individual herd heritabilities could be assumed to be homogeneous (through lack of accurate estimates thereof), and that individual herd phenotypic variances could be estimated and regressed to a prior estimate, depending on the number of records per herd.

The difference between variance component estimates for yield traits in lactation 2 from a bivariate and univariate analysis was most likely due to culling of heifers being dependent on their first lactation performance (chapter 5). Since the exact culling process is unknown, estimating variances bivariately may give biased variance components for (co)variance in later lactations. In particular the (co)variances between yield traits within lactation 2 and 3, and the covariances between lactations 2 and 3 are expected to be biased. A general multivariate analysis using all traits in all lactations would be appropriate to investigate potential selection bias, but this is computationally not (yet) feasible. Still, despite possible biases in parameter estimates, the proportionality of the (co)variance matrices in different lactations is striking, and small changes in estimates would not change the observed proportionality.

In chapter 6 only first order approximations are given for loss in efficiency when simplified covariance structures are assumed for prediction of breeding values. Clearly, more research is needed to investigate the loss in genetic progress when such assumptions are made. A related problem, particularly relevant to the breeding industry, is that of estimating genetic trend from BLUP analyses. The robustness of animal model trend estimates to different models and parameters is not fully understood, and needs further research.

Using the parameter estimates from chapter 5 for milk, fat and protein yield in lactations 1-3, it was found that a repeatability model of 3 new traits, created by applying the canonical transformation for milk, fat and protein yield in lactation 1 to

yield traits within lactations 2 and 3, was highly efficient. This conclusion was based on index selection calculations. In practice, selection acts on an unknown combinations of yield traits in several lactations, and a repeatability model on the separate yield traits is likely to give biased predictions of breeding values, since correlations between yield traits, and hence selection bias, are not taken into account properly. It is concluded, therefore, that a repeatability model on the canonical variates is more efficient and should be preferred to the (standard) separate evaluation of milk, fat and protein yield.

REFERENCES

Anderson, T.W. (1958). *An introduction to multivariate statistical analysis*. John Wiley & Sons, New York.

Bartlett, M.S. (1937). Properties of sufficiency and statistical tests. *Proc. Roy. Soc. Series* **B**, **160**:268-282.

Beaumont, C. (1988). Restricted Maximum Likelihood estimation of genetic parameters for the first three lactations in the Montbeliarde breed. *Abstracts from the 14th International Biometric Conference*: C–13.6, Namur, Belgium.

Boldman, K.G. and Freeman, A.E. (1988). Estimates of genetic and environmental variances of first and later lactations at different production levels. *J. Dairy Sci.* **71** (suppl. 2):81-82.

Boldman, K.G. and Freeman, A.E. (1990). Adjustment for heterogeneity of variances by herd production level in the dairy cow and sire evaluations. *J. Dairy Sci.* **73**:503-512.

Bonaiti, B., and Boichard, D. (1990). Benefits from animal model evaluation of dairy cattle in France. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.*, Edinburgh, **Vol. XIII**:364-373.

Brotherstone, S. and Hill, W.G. (1986). Heterogeneity of variance amongst herds for milk production. *Anim. Prod.* **42**:297-303.

Brotherstone, S., and Hill, W.G. (1991a). Dairy herd life in relation to linear type traits and production. 1. Phenotypic and genetic analyses in pedigree type classified herds. *Anim. Prod.* (in press).

Brotherstone, S., and Hill, W.G. (1991b). Dairy herd life in relation to linear type traits and production. 2. Genetic analyses for pedigree and non-pedigree cows. *Anim. Prod.* (in press).

Brotherstone, S., Hill, W.G. and Thompson, R. (1989). Components of environmental variance of yields of cows in different lactations. *Anim. Prod.* **48**:283-291.

Bulmer, M.G. (1971). The effect of selection on genetic variability. *American Naturalist* **105**:201-211.

Bulmer, M.G. (1980). *The mathematical theory of quantitative genetics*. Oxford University Press, Oxford, U.K.

Carabaño, M.J., Wade, K.M., and Van Vleck, L.D. (1990). Genotype by environment interaction for milk and fat production across regions of the United States. *J. Dairy Sci.* **73**:173-180.

Chang, H.L., Fernando, R.L., and Gianola, D. (1989). Inverse of an additive x additive relationship matrix due to sires and maternal grandsires. *J. Dairy Sci.* **72**:3023-3034.

Chauhan, V.P.S. (1987a). Dairy sire evaluation fitting some of the

herd-year-seasons effects as random. *Livest. Prod. Sci.* **16**:117-130.

Chauhan, V.P.S. (1987b). Partitioning of herd, year and season variation in milk production. *Livest. Prod. Sci.* **16**:107-116.

Chauhan, V.P.S. and Thompson, R. (1986). Dairy sire evaluation using a "rolling months" model. *J. Anim. Breedg. Genet.* **103**:321-333.

Chesnais, J.P. and Song, H. (1988). Estimation of genetic value of beef cattle from on-farm records using the animal model. *In: Proc. Third World Congress on sheep and beef cattle breeding*, Paris **Vol. 1**:219-237.

Dempfle, L. and Grundl, E. (1988). Identification of superior animals and their use in improvement. *In: Advances in Animal Breeding: proceeding of the world symposium in honour of professor R.D. Politiek* :56-72.

Dickerson, G.E. (1962). Implication of Genetic-Environmental interaction in Animal Breeding. *Anim. Prod.* **4**:47-63.

Dong, M.C. and Mao, I.L. (1990). Heterogeneity of (co)variance and heritability in different levels of intraherd milk production variance and herd average. *J. Dairy Sci.* **73**:843-851.

Ducrocq, V. and Quaas, R.L. (1988). Prediction of genetic response to truncation selection across generations. *J. Dairy Sci.* **71**:2543-2553.

Ducrocq, V., Boichard, D., Bonaiti, B., Barbat, A. and Briend, M. (1990). A pseudo-absorption strategy for solving animal model equations for large data files. *J. Dairy Sci.* **73**:1945-1955.

Falconer, D.S. (1952). The problem of environment and selection. *American Naturalist* **86**:293-298.

Falconer, D.S. (1983). *Introduction to quantitative genetics. 2nd Edition.* Longman, London.

Falconer, D.S. (1990). Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. *Genet. Res. Camb.* **56**:57-70.

Famula, T.R. (1989). Detection of heterogeneous variance in herd production groups. *J. Dairy Sci.* **72**:715-721.

Fernando, R.L. and Gianola, D. (1990). Statistical inferences in populations undergoing selection or non-random mating. *In: Advances in statistical methods for genetic improvement of livestock*: 437-453. Gianola, D. and Hammond, K. (ed.). Springer-Verlag, Berlin.

Fisher, R.A. 1921. On the "probable error" of a coefficient of correlation deduced from a small sample. *Metron* **1 part 4**:1-32.

Fisher, R.A. (1956). *Statistical Methods and Scientific Inference.* Oliver and Boyd, Edinburgh-London.

Foulley, J.L. and Henderson, C.R. (1989). A simple model to deal with sire by treatment interactions when sires are related. *J. Dairy Sci.* **72**:167-172.

Foulley, J.L., Gianola, D., San Cristobal, M., and Im, S. (1990). A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *J. Dairy Sci.* **73**:1612-1624.

Gianola, D. (1986). On selection criteria and estimation of parameters when the variance is heterogeneous. *Theor. Appl. Genet.* **72**:671-677.

Gianola, D. (1990). Can BLUP and REML be improved upon? *Proc. 4th World Congr. Genet. Appl. Livest. Prod.*, Edinburgh, **Vol. XIV**:445-449.

Gianola, D., Im, S. and Fernando, R.L. (1988). Prediction of breeding values under Henderson's selection model: a revisitation. *J. Dairy Sci.* **71**:2790-2798.

Goddard, M. (1990). Discussion summary on Part IV: Selection and non-random mating. *In: Advances in statistical methods for genetic improvement of livestock*: 474-475. Gianola, D. and Hammond, K. (ed.). Springer-Verlag, Berlin.

Graser, H-U, Smith, S.P. and Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by Restricted Maximum Likelihood. *J. Anim. Sci.* **64**:1362-1370.

Griffing, B. (1960). Theoretical consequences of truncation selection based on the individual phenotype. *Austr. J. Biol. Sci.* **13**:307-343.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* **72**:320-340.

Hayes, J.F. and Hill, W.G. (1980). A reparameterisation of a genetic index to locate its sampling properties. *Biometrics* **36**:237-248.

Henderson, C.R. (1973). Sire evaluation and genetic trends. *In: Proc. of the Animal Breeding and Genetics Symposium in honor of Dr. J. Lush*, 10-41. American Society of Animal Science, American Dairy Science Association, American Poultry Science Association, Champaign, Illinois.

Henderson, C.R. (1975a). Best Linear Unbiased Estimation and Prediction under a selection model. *Biometrics* **31**:423-447.

Henderson, C.R. (1975b). Comparison of alternative sire evaluation methods. *J. Anim. Sci.* **41**:760-770.

Henderson, C.R. (1975c). Use of all relatives in intra-herd predictions of breeding values and production abilities. *J. Dairy Sci.* **58**:1910-1921.

Henderson, C.R. (1976). A simple method of computing the inverse of a numerator relationship matrix used in prediction of breeding values.

Biometrics **32**:69-83.

Henderson, C.R. (1988). Theoretical basis and computational methods for a number of different Animal Models.*J. Dairy Sci.* **71** (suppl. 2):1-16.

Henderson, C.R. (1990). Accounting for selection and mating biases in genetic evaluations. *In: Advances in statistical methods for genetic improvement of livestock*: 413-436. Gianola, D. and Hammond, K. (ed.). Springer-Verlag, Berlin.

Hill, W.G. (1984). On selection among groups with heterogeneous variance. *Anim. Prod.* **39**:473-477.

Hill, W.G. and Swanson, G.J.T. (1983). A selection index for dairy cows. *Anim. Prod.* **37**:313-319.

Hill, W.G., Edwards, M.R., Ahmed, M.-K.A. and Thompson, R. (1983). Heritability of milk yield and composition at different levels and variability of production. *Anim. Prod.* **36**:59-68.

Interbull (1986). *Procedures for international comparisons of dairy sires - current practice and evaluation of methods*. Bulletin no. **1**, Uppsala, Sweden.

Interbull (1988). *Sire evaluation procedures for dairy production traits practised in various counties*. Bulletin no. **3**, Uppsala, Sweden.

Jones, L.P., and Goddard, M.E. (1990). Five years experience with the animal model for dairy cattle evaluations in Australia. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.*, Edinburgh, **Vol. XIII**:382-385.

Kackar, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for linear models. *Commun. Statist.-Theor. Meth.* **A10**:1249-1261.

Kendall, M.G. and Stuart A. (1973). *The advanced theory of statistics*. 3rd edition. **Vol. 2**: Inference and relationship. Griffin, London, U.K.

Kennedy, B.W., Schaeffer, L.R. and Sorensen, D.A. (1988). Genetic properties of Animal Models. *J. Dairy Sci.* **71** (suppl. 2):17-26.

Lawley, D.N. and Maxwell, A.E. (1971). *Factor analysis as a statistical method*. 2nd edition. Butterworth & Co, London, U.K.

Lofgren, D.L., Vinson, W.E., Pearson, R.E. and Powell, R.L. (1985). Heritability of milk yield at different herd means and variances for production. *J. Dairy Sci.* **68**:2737-2739.

Maijala, K. and Hanna, M. (1974). Reliable phenotypic and genetic parameters in dairy cattle. *In: Proc. First World Congress of genetics applied to livestock production*, Madrid **Vol. 1**:541-563.

Meinert, T.R., Pearson, R.E., Vinson, W.E. and Cassell, B.G. (1988). Prediction of daughter's performance from dam's cow index adjusted for within-herd variance. *J. Dairy Sci.* **71**:2220-2231.

Meyer, K. (1983). Scope for evaluating dairy sires using first and second lactation records. *Livest. Prod. Sci.* **10**:531-553.

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

Meyer, K. (1985). Maximum Likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* **41**:153-165.

Meyer, K. (1987). Estimates of variance due to sire*herd interaction and environmental covariances between paternal half-sibs for first lactation dairy production. *Livest. Prod. Sci.* **17**:95-115.

Meyer, K. (1989). Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet. Sel. Evol.* **21**:317-340.

Meyer, K. (1991). Estimating variances and covariances for multivariate animal models by restricted maximum likelihood. *Genet. Sel. Evol.* **23**:67-83.

Meyer, K. and Thompson, R. (1984). Bias in variance and covariance component estimators due to selection on a correlated trait. *J. Anim. Breedg. Genet.* **101**:33-50.

Mirande, S.L. and Van Vleck, L.D. (1985). Trends in genetic and phenotypic variances for milk production. *J. Dairy Sci.* **68**:2278-2286.

Mood, A.M., Graybill, F.A., and Boes, D.C. (1973). *Introduction to the theory of statistics*. McGraw-Hill, New York.

Norman, H.D. (1974). Factors that should be considered in a national sire summary model. *J. Dairy Sci.* **57**:955-962.

Norman, H.D., McDaniel, B.T. and Dickinson, F.N. (1972). Modified herdmate comparison procedure that weights deviation yields considering daughters' distribution across herds. *J. Dairy Sci.* **55** (abstr.):691.

Norman, H.D., Powell, R.L. and Wright, J.R. (1985). Changes of evaluation for Natural Service sampled bulls brought into artificial insemination. *J. Dairy Sci.* **68**:1513-1522.

Oldenbroek, J.K. (1988). *Feed intake and energy utilization in dairy cows of different breeds*. PhD-thesis, Wageningen, Holland.

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

Pearson, R.E., Meinert, T.R. and Smothers, C.D. (1988). Response to selection for milk and type in herds varying in herd mean and within-herd variance. *J. Dairy Sci.* **71** (suppl. 2):79-80.

Quaas, R.L. (1976). Computing the diagonal elements of a large numerator relationship matrix. *Biometrics* **32**:949-953.

Quaas, R.L., Garrick, D.J., and McElhenney, W.H. (1989). Multiple trait prediction for a type of model with heterogeneous genetic and residual covariance structures. *J. Anim. Sci.* **67**:2529-2535.

Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics* **15**:469-485.

Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Anim. Prod.* **8**:95-108.

Robertson, A. and Hill, W.G. (1983). Population and quantitative genetics of many linked loci in finite populations. *Proc. R. Soc. Lond. Series* **B**, **219**:253-264.

Robinson, J.A.B. and Chesnais, J.P. (1988). Application of the Animal Model on a national basis to the evaluation of Canadian livestock. *J. Dairy Sci.* **71** (suppl. 2):70-78.

Sales, J. and Hill, W.G. (1976a). Effect of sampling errors on efficiency of selection indices. 1.: Use of information from relatives for single trait improvement. *Anim. Prod.* **22**:1-17.

Sales, J. and Hill, W.G. (1976b). Effect of sampling errors on efficiency of selection indices. 2.: Use of information on associated traits for improvement of a single important trait. *Anim. Prod.* **23**:1-14.

San Cristobal, M., Foulley, J.L., and Gianola, D. (1990). Inference about heterogeneous U-components of variance in mixed linear models. *41st EAAP meeting.* Toulouse, France.

Schaeffer, L.R., Wilton, J.W. and Thompson, R. (1978). Simultaneous estimation of variance and covariance components from multitrait mixed model equation. *Biometrics* **34**:199-208.

Schaeffer, L.R., Kennedy, B.W., and Gibson, J.P. (1989). The inverse of the gametic relationship matrix. *J. Dairy Sci.* **72**:1266-1272.

Searle, S.R. (1966). *Matrix algebra for the biological sciences.* John Wiley & Sons, New York.

Searle, S.R. (1979). *Notes on variance component estimation.* Paper **BU-673-M** of Biometrics Unit, Cornell University, U.S.A.

Short, T.H., Blake, R.W., Quaas, R.L., and Van Vleck, L.D. (1990). Heterogeneous within-herd variance. 1. Genetic parameters for first and second lactation milk yield of grade Holstein cows. *J. Dairy Sci.* **73**:3312-3320.

Smith, S.P. and Graser, H-U. (1986). Estimating variance components in a class of mixed models by Restricted Maximum Likelihood. *J. Dairy Sci.* **69**:1156-1165.

Smith, S.P. and Maki-Tanila, A. (1990). Genotypic covariance matrices and their inverses for models allowing dominance and inbreeding. *Genet. Sel. Evol.* **22**:65-91.

Sorensen, D.A. (1989). Notes for the Pig Breeders Round Table, Wye College, U.K.

Sorensen, D.A. and Kennedy, B.W. (1983). The use of the relationship matrix to account for genetic drift variance in the analysis of genetic experiments. *Theor. Appl. Genet.* **66**:217-220.

Sorensen, D.A. and Kennedy, B.W. (1984). Estimation of response to selection using Least-Squares and Mixed Model methodology. *J. Anim. Sci.* **58**:1097-1106.

Swalve, H. and Van Vleck, L.D. (1987). Estimation of genetic (co)variances for milk yield in first three lactations using an animal model and restricted maximum likelihood. *J. Dairy Sci.* **70**:842-849.

Swanson, G.J.T. (1991). Organisation of the dairy breeding programme in the United Kingdom. *Proc. meeting "Dairy cattle breeding in Europe"*, Santander, Spain.

Swanson, G.J.T. and Gnanasakthy, A. (1991). Estimation of genetic parameters for milk cell count and correlations with production traits in Friesian Holstein heifers. *Proc. winter meeting Brit. Soc. Anim. Prod.* (abstr.): paper **168**.

Thompson, R. (1977). Estimation of quantitative genetic parameters. *In: Proc. Int. Conf. on Quantitative Genetics*: 639-657, Iowa State Univ. Press, Ames.

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

Thompson, R. and Meyer, K. (1986a). A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livest. Prod. Sci.* **15**:299-313.

Thompson, R. and Meyer, K. (1986b). Estimation of variance components: What is missing in the EM algorithm? *J. Statis. Comput. Simul.* **24**:215-230.

Thompson, R. and Hill, W.G. (1990). Univariate REML analyses for multivariate data with the animal model. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.*, Edinburgh, **Vol. XIII**:484-487.

Turelli, M. and Barton, N.H. (1990). Dynamics of polygenic characters under selection. *Theor. Pop. Biol.* **38**:1-57.

Van der Werf, J.H.J., and De Boer, W. (1989a). Influence of non-additive effects on estimation of genetic parameters in dairy

cattle. *J. Dairy Sci.* **72**:2606-2614.

Van der Werf, J.H.J., and De Boer, W. (1989b). Estimation of genetic parameters in a crossbred population of black and white dairy cattle. *J. Dairy Sci.* **72**:2615-2623.

Van Vleck, L.D. (1986). Evaluation of dairy cattle breeding programs: specialised milk production. *In: Proc. Third World Congress on Genetics applied to livestock production*, Lincoln, Nebraska **Vol. 9**:141-152.

Van Vleck, L.D. (1988a). Selection when traits have different genetic and phenotypic variances in different environments. *J. Dairy Sci.* **70**:337-344.

Van Vleck, L.D. and Dong, M.C. (1988). Genetic (co)variances for milk, fat, and protein yield in Holsteins using an animal model. *J. Dairy Sci.* **71**:3040-3046.

Van Vleck, L.D., Dong, M.C., and Wiggans, G.R. (1988). Genetic (co)variances for milk and fat yield in California, New York, and Wisconsin for an animal model by restricted maximum likelihood. *J. Dairy Sci.* **71**:3053-3060.

Vinson, W.E. (1987). Potential bias in genetic evaluations from differences in variation within herds. *J. Dairy Sci.* **70**:2450-2455.

Visscher, P.M., and Thompson, R. (1990). REML estimates of parameters for fat yield in pedigree herds in the U.K. using an individual animal model; Male and female heritability estimates. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.* **Vol XIV**:233-236.

Wade, K.M., Quaas, R.L., and Van Vleck, L.D. (1990). Mixed linear models with an autoregressive error structure. *In: Proc. 4th World Congress on Genet. Appl. Livest. Prod.*, Edinburgh, **Vol. XIII**:508-511.

Weller, J.I. (1988). Inclusion of partial lactations in the genetic analysis of yield traits by differential weighting of records. *J. Dairy Sci.* **71**:1873-1879.

Westell, R.A., Quaas, R.L., and Van Vleck, L.D. (1988). Genetic groups in an animal model. *J. Dairy Sci.* **71**:1310-1318.

Wiggans, G.R., Misztal, I. and Van Vleck, L.D. (1988a). Implementation of an Animal Model for genetic evaluation of dairy cattle in the United States. *J. Dairy Sci.* **71** (suppl. 2):54-69.

Wiggans, G.R., Misztal, I. and Van Vleck, L.D. (1988b). Animal Model evaluation of Ayrshire milk yield with all lactations, herd-sire interaction, and groups based on unknown parents. *J. Dairy Sci.* **71**:1319-1329.

Wilmink, J.B.M. (1988). Effects of incomplete records on relations among cumulative yields in first lactation and on extension of part lactations. *Livest. Prod. Sci.* **18**:19-34.

Winkelman, A. and Schaeffer, L.R. (1988). Heterogeneity of variances among herds and its effects on dairy sire evaluation. *J. Dairy Sci.* **71** (suppl. 2):84.

Wray, N.R. and Hill, W.G. (1989). Asymptotic rates of response from index selection. *Anim. Prod.* **49**:217-227.

APPENDIX

Published papers:

- Visscher, P.M. (1990). REML estimates of parameters for fat yield in pedigree herds in the U.K. using an individual animal model; Individual herd analyses vs. a combined analysis. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.* **XIV**:229-232.

- Visscher, P.M. and Thompson, R. (1990). REML estimates of parameters for fat yield in pedigree herds in the U.K. using an individual animal model; Male and female heritability estimates. *In: Proc. 4th World Congr. Genet. Appl. Livest. Prod.* **XIV**:233-236.

The follow pages were poor originals from the HardCopy Theses

REML ESTIMATES OF PARAMETERS FOR FAT YIELD IN PEDIGREE HERDS IN THE U.K. USING AN INDIVIDUAL ANIMAL MODEL: INDIVIDUAL HERD ANALYSES VS. A COMBINED ANALYSIS

P.M. Visscher
Department of Genetics, University of Edinburgh,
West Mains Road, Edinburgh EH9 3JN, U.K.

SUMMARY

Restricted Maximum Likelihood (REML) parameters for fat yield were estimated in large Holstein-Friesian pedigree herds, using an Individual Animal Model (IAM). Heritability estimates from individual herds were similar, but the genetic and environmental variances differed among herds.

INTRODUCTION

In the 1990's the IAM (Individual Animal Model) is likely to become the model of evaluation for dairy cattle. Some countries already have implemented their IAM evaluation (Wiggans et al, 1988), others will follow within the next few years. The IAM requires fewer assumptions about the data than a sire model, for example random non-mating of males and females is taken into account. However, the IAM may be susceptible to problems which previously were of lesser importance for evaluation with a BLUP sire model. One such a problem is that of heterogeneity of variance, i.e. the variation of EBV's (Estimated Breeding Values) among cows within a herd or herd-year-season is influenced by the phenotypic variation within that environment. If it is not known whether the genetic variance, the environmental variance, or both variances are heterogeneous, then the effect on accuracy of selection is not predictable. Previously, if young bulls were tested among herds from many different mean and variance groups and evaluated with a sire model, their EBV was unlikely to be heavily influenced by heterogeneity of variance between herds or herd-year-seasons. For the problem of het# of variance it is not clear what the effect on EBV's will be using an IAM.

Usually the estimates of variances (or their ratios) required for BLUP are derived from REML (Restricted Maximum Likelihood; Patterson and Thompson, 1971) procedures using a similar model of analysis to that used to predict breeding values. For U.K. data the use of an IAM for estimating population parameters had not been investigated. The aim of this study was to calculate REML estimates using an IAM, with special attention to the problem of heterogeneity of within herd variance. To try to explain potential heterogeneity of variance the data were analysed initially at herd level. The estimations were carried out using a REML program written by Karin Meyer (Meyer, 1989).

MATERIAL AND METHODS

A sample of 26 large Holstein Friesian (HF) pedigree herds was taken. Before editing 7979 first lactation fat yield records were present, from cows calving between 1981 and 1986. The overall mean and (uncorrected) standard deviation were 213 ± 43.9 kg fat. The ranges of herd means and herd standard deviations were 170.3-263.6 and 25.0-48.9 respectively. The correlation between herd mean and standard deviation was close to zero (+ 0.02). The average North American HF percentage of the cows was 23 %. 581 sires were represented in the complete data set, both young and old (proven) sires. 186 bulls only had 1 daughter, whereas proven bulls had up to 450 daughters present. 1740 daughter-dam pairs with records were present, of which only 6 pairs were not in the same herd. After editing 7720 records (from 574 sires) were left.

The assumed linear model, with one random effect besides the residual effect, was:
$y = Xb + Zu + e$ and:
$v(y) = ZAZ'\sigma_a^2 + I\sigma_e^2 = ZGZ' + R$; with the usual definitions:

y,b,u are vectors of the observations, fixed effects and individual animal effects respectively, X,Z are the known incidence matrices for the fixed and random effects, and A is the numerator relationship matrix.

HYS were the only fixed effects, and age at calving, percentage HF and lactation length were fitted as covariables. The (natural) Log-Likelihood (L) for a model with one other random effect besides the residual component is (e.g. Harville, 1977; Searle, 1979):

$$L = -1/2\{ \log|R| + \log|A| + \log|C| - \log|X'X| + y'Py \}$$

Where C= full rank submatrix of the coefficient matrix and y'Py= residual Sum of Squares, with P a projection matrix. Significant tests for heritability estimates were carried out as Likelihood ratio tests (see e.g. Mood et al, 1973). Standard errors on parameter estimates were obtained through approximating the likelihood curve by a quadratic function, and taking the second differential with respect to the parameter of interest.

Three different models were fitted. In analysis Ia herds were evaluated separately and, assuming independence and equal weightings, maximum (log)likelihoods were summed and heritabilities pertaining to those maxima were averaged over all 26 estimates. The standard error presented for Ia is from the empirical variance of the estimates. Ib uses the same 26 estimates, but now the overall maximum and likelihood is obtained by weighting the estimates according to the amount of information present in each data set. This was investigated by summing up all 26 likelihood curves and fitting a quadratic to the obtained curve to obtain the maximum likelihood estimate. In analysis II the same data were used, but all herds were combined in one data set to give the overall heritability estimate and likelihood.

RESULTS

Table 1 gives a summary of the results of the 3 different analyses. Some summary statistics for the individual herd estimates are presented in table 2.

Table 1: Comparison of different REML evaluations

	Ia	Ib	II
L	-27239.7	-27251.4	-27998.6
h^2	0.388	0.387	0.379
s.e.(h^2)	0.037	0.033	0.037
$\Psi(h^2)$	731	918	734

s.e.= standard error
$\Psi(h^2)$= information for h^2 estimate= $\{var(h^2)\}^{-1}$
L= log(Likelihood)
Analyses: Ia: individual herds
Ib: combined estimates of individual herds
II= all herds together

The combined analysis II shows a lower likelihood than the summed maximum likelihoods from Ia, which is not surprising given that II was analysed with fewer degrees of freedom i.e. the separate herd analyses allow for more parameters to be fitted; implicitly a sire by herd interaction, a herd by genetic variance interaction and a herd by residual variance interaction

were fitted in analysis Ia. The heritability estimate is hardly different; this can be a result of the particular design, or may be due to a large contribution of daughter-dam comparisons. As expected the s.e. is slightly larger than in Ib, confirming that the separate data sets were not independent, since there is a positive covariance among individual herd heritability estimates.

Table 2: Summary statistics for individual herd estimates

PARAMETER	MEAN	MIN	MAX	Q1	Q3	STDEV
"Raw" σ_p^2	1247.1	625.0	2391.2	967.5	1532.8	411.1
REML results:						
σ_p^2	834.5	482.7	1353.4	642.0	943.3	225.3
σ_e^2	330.6	17.4	771.1	192.0	513.2	184.4
σ_e^2	504.0	161.6	839.1	362.5	619.9	186.1
h^2	0.388	0.030	0.800	0.273	0.504	0.186

The statistics are respectively: mean, minimum, maximum, lower quartile, upper quartile and the empirical standard deviation.
Raw σ_p^2: phenotypic variance before any corrections.

Testing each herd heritability from I against the overall heritability from II showed no significant difference at the 1 % level. A single likelihood ratio test, comparing the likelihood from Ia (= sum all maxima) with Ib (= the maximum likelihood for combining the 26 estimates to one estimate) also showed no significant difference between the heritability estimates (-2*log(difference likelihood)= 23.4 for 25 degrees of freedom). The overall heritability agrees well with the most recent U.K. estimate for pedigree herds (Meyer, 1987).

In further likelihood ratio tests it was assumed that the heritability was the same for all herds, allowing for simple tests for the variances. This assumption results in equivalent tests for genetic and environmental variances. The test for variances resulted in 11 out of 26 estimates differing from an overall variance estimate. A single likelihood ratio test also showed significance at the 1 % level (-2*log(difference)= 44.2 with 25 degrees of freedom). Clearly a relatively large heterogeneity of variance among herds is present for this data set. The estimated phenotypic variances for the separate herds showed a nearly three-fold difference between the lowest and highest phenotypic variance.

DISCUSSION

The sample of the pedigree herds may not be representative for all the pedigree herds or for the non-pedigree herds of the black-and-white breed. Meyer (1987), fitting a sire model, found substantially higher heritabilities for fat yield in pedigree herds compared with non-pedigree herds. Furthermore, the amount of information per herd may not be sufficient to detect real differences in heritabilities and variances. More information per herd may also provide sufficient power to distinguish between heterogeneity of genetic and heterogeneity of environmental variance. However, despite the large standard errors for heritability estimates from individual herds (approximately 0.19), few extreme heritability estimates were obtained in this study and the pedigree herds were the largest herds available.

For these data, heritabilities seemed to be the same for all herds. Other authors also have found higher genetic variances in the more variable herds for production traits, sometimes by finding higher heritabilities in those herds (e.g. Hill et al, 1983; Lofgren et al, 1985; Boldman and Freeman, 1988). The observed heterogeneity of variance could not be explained by a scale effect, since the correlation between herd mean and variance was close to zero. Assuming equal heritabilities across herds makes it relatively easy to take heterogeneity of variance into

account in a BLUP-evaluation: either the data can be precorrected for the within-herd phenotypic variance, or the diagonal elements for animals and fixed effects can be manipulated during the evaluation. Estimating phenotypic variances from small herds may cause sampling problems. Regressing estimates from individual herds to some overall estimate of the phenotypic variance, where the regression coefficient depends on the herd size (degrees of freedom), was discussed by Brotherstone and Hill (1986).

If further investigation indicates that heritabilities are not the same for all herds, then a different approach should be taken. A multi-trait approach seems theoretically best (see e.g. Gianola, 1986), but it may be tedious to estimate genetic and phenotypic parameters for all herds in order to group them according to some function of the estimated parameters.

ACKNOWLEDGEMENTS

Financial support from the Milk Marketing Boards in the U.K. is appreciated. I wish to thank Robin Thompson, Bill Hill and Naomi Wray for many helpful comments.

REFERENCES

BOLDMAN, K.G. and FREEMAN, A.E. 1988. J. Dairy Sci. 71 suppl. 2:81-82.
BROTHERSTONE, S. and HILL, W.G. 1986. Anim. Prod. 42:297-303.
GIANOLA, D. 1986. Theor. Appl. Genet. 72:671-677.
HARVILLE, D.A. 1977. J. Am. Stat. Assoc. 72:320-340.
HILL, W.G., EDWARDS, M.R., AHMED, M.-K.A. and THOMPSON, R. 1983. Anim. Prod. 36:59-68.
LOFGREN, D.L., VINSON, W.E., PEARSON, R.E. and POWELL, R.L. 1985. J. Dairy Sci. 68:2737-2739.
MEYER, K. 1987. Livest. Prod. Sci. 17:95-115.
MEYER, K. 1989. Genet. Sel. Evol. 21:317-340.
MOOD, A.M., GRAYBILL, F.A., and BOES, D.C. 1973. McGraw-Hill, New York.
PATTERSON, H.D. and THOMPSON, R. 1971. Biometrika 58:545-554.
SEARLE, S.R. 1979. Paper BU-673-M of Biometrics Unit, Cornell University, U.S..
WIGGANS, G.R., MISZTAL, I. and VAN VLECK, L.D. 1988. J. Dairy Sci. 71 suppl. 2:54-69.

REML ESTIMATES OF PARAMETERS FOR FAT YIELD IN PEDIGREE HERDS IN THE U.K. USING AN INDIVIDUAL ANIMAL MODEL; MALE AND FEMALE HERITABILITY ESTIMATES

P.M. Visscher and R. Thompson
Department of Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, U.K. and AFRC Institute of Animal Physiology and Genetics Research, Roslin EH25 9PS, Midlothian, U.K.

SUMMARY

In dairy cattle, estimates of heritability from daughter-dam regression are usually higher than from paternal half-sib covariance, and the genetic variance of bulls is less than that of females. These phenomena are modelled by introducing an extended IAM (Individual Animal Model) with "male" (h_m^2) and "female" (h_f^2) heritabilities. Estimates of h_m^2 and h_f^2 were 0.280 ± 0.045 and 0.477 ± 0.052 from a data set of 26 pedigree Holstein-Friesian herds. It is shown that a quadratic approximation of the likelihood surface for the two heritabilities is insufficient, both for the data set and for a hierarchical balanced mating design.

INTRODUCTION

The genetic and phenotypic parameters required for the BLUP evaluations in dairy cattle are usually estimated using a similar model of analysis to that used to predict breeding values. In the 1980's that model usually was a sire model, and the estimation procedure which has become widely adopted is REML (Restricted Maximum Likelihood; Patterson and Thompson, 1971). In the present decade the IAM (Individual Animal Model) is likely to become the model of evaluation (Wiggans et al, 1988), resulting in a joint evaluation of males and females.

This study introduces a model to investigate two related phenomena. Firstly it is well known that heritability estimates from daughter-dam regression usually are found to be higher than heritability estimates from PHS (paternal half-sib) correlation. Secondly it is likely that the sire genetic variance is less than the genetic variance in females, because bull dams and sires are selected more intensely than cow dams and sires. In estimating parameters from an IAM one is constrained by the size of the data set to be analysed, and one suggestion is to analyse a small number of herds. In this type of data several assumptions could be made about the genetic variances of sires of animals. With this type of data structure, selection in females may be taken account of by ML methods, but it is most unlikely that selection in males can be accounted for. We extend the IAM to take account of these two phenomena by introducing and estimating "male" (h_m^2) and "female" (h_f^2) heritabilities.

MATERIAL AND METHODS

A description of the data set was given by Visscher (1990). All analyses presented here were carried out using the complete data set of 7720 records and 12620 animals. Analysis I is the standard IAM estimation, also presented previously (Visscher, 1990).

In analysis II two random components (besides the residual component) were estimated; a female heritability was estimated by fitting a random effect for females, only including female relationships in the covariance matrix, and a male heritability was estimated by fitting sires as an uncorrelated random effect with an identity covariance matrix. For analysis III an extended IAM was constructed by partitioning the relationship matrix A into male and female contributions. Let subscript *f* and *m* denote female and male respectively. Writing the relationship matrix as the product of a lower triangular, a diagonal and an upper triangular matrix (Thompson, 1977), gives $A=TDT'$. Now partition D into a male and female part, giving:

$$D = \begin{bmatrix} D_m & 0 \\ 0 & D_f \end{bmatrix} \quad \text{then: } A = T \begin{bmatrix} D_m & 0 \\ 0 & 0 \end{bmatrix} T' + T \begin{bmatrix} 0 & 0 \\ 0 & D_f \end{bmatrix} T'$$

The variance among breeding values can be written as: $v(u) = TD_m'T'\sigma_{Am}^2 + TD_f'T'\sigma_{Af}^2$, where D_m' and D_f' are diagonal matrices with the appropriate number of zero diagonal elements for females and males respectively. The genetic variance in females thus has an individual component and a male and female parent component. Essentially this model allows the sires to come from a population with genetic variance σ_{Am}^2. Using the above model, the phenotypic variance was estimated as:
$\sigma_p^2 = 3/4*\sigma_{Af}^2 + 1/4*\sigma_{Am}^2 + \sigma_e^2$,
since records were only on females. The male and female heritabilities presented are the ratio of the two genetic variances to this estimated phenotypic variance. Under this model the daughter-dam covariance is $0.5*\sigma_{Af}^2$ and the paternal half-sib covariance is $0.25*\sigma_{Am}^2$. One would expect these relationships to provide most of the information on genetic variance in the data.

Model IV is a sire model, for which all 574 sires were fitted as base sires, i.e. no relationships between sires were fitted. Using the results from analyses III and II an attempt was made to predict the IAM results from I by weighting the two heritability estimates:
$h_{mf}^2 = w_m h_m^2 + w_f h_f^2$, with the vector of weights, w, calculated as $w = \Psi(1'\Psi 1)^{-1}1$,
where Ψ is the information matrix and 1 is a vector of ones.

RESULTS

The results from the 4 analyses are presented in table 1.

Table 1: Comparison of different REML evaluations

		I	II	III	IV
	L	-27998.6	-27997.9	-27995.4	-28043.8
	h_f^2		0.408	0.477	
	s.e.(h_f^2)		0.043	0.052	
	$\Psi(h_f^2)$		542	370	
	h_m^2		0.288	0.280	0.299
	s.e.(h_m^2)		0.046	0.045	0.048
	$\Psi(h_m^2)$		490	509	430
Combined	h_{mf}^2	0.379	0.351	0.365	
	$\Psi(h_{mf}^2)$	734	1172	1001	
Estimate	$r(h_f^2,h_m^2)$		-0.14	-0.14	

L= log(Likelihood); s.e.= standard error; $\Psi(h^2)$= information on h^2 estimates
$r(h_f^2,h_m^2)$= correlation between estimates derived from the Ψ-matrix
h_{mf}^2= combined IAM estimate
Combined estimate: using the estimates and curvature at the maxima

Model I was discussed by Visscher (1990). Model II allows for heterogeneity of genetic variance to some extent, but the covariance structure is only an approximation of the structure in analysis I of III. For example, sires are not linked to their grand-offspring in analysis II. The estimate of the male heritability in II (4 times the intra-class correlation) is close to the heritability estimate using a sire model (analysis IV). The contribution from males and females

seems similar, but the predicted heritability and curvature for the combined (IAM) estimate are not very close to the observed values in I. The likelihood in II is higher than in I, but not significantly so at the 10% level. Analysis III clearly fits the data best. The likelihood difference with I is significant at the 1% level (for 1 degree of freedom). The female heritability is substantially higher than the male heritability. Again the simple weighting of the estimates did not result in the values from I. Comparing II with III the information on the female heritability was reduced for III, while the male heritability remained nearly constant.

In an attempt to explain the difference between the prediction from III and the observed values from I, quadratic functions in h_m^2 and h_f^2 were fitted to various grids of (h_m^2,h_f^2) values. The first part of table 2 shows the second differentials of log-likelihood with respect to the heritabilities, or curvature matrix, at several grid values. Inverting this curvature matrix at the REML heritability estimates gives the asymptotic covariance matrix. Each grid consisted of 9 equally spaced points around the presented heritability values.

Table 2: Curvature of log-likelihood for various values of heritability estimates from the data set and from a hierarchical balanced design

		CURVATURE FROM DATA				CURVATURE FROM HIERARCHICAL DESIGN		
		h_m^2				h_m^2		
		0.280	0.335	0.379		0.280	0.330	0.379
		362	373	383	$\Psi(h_f^2)$	280	290	300
	0.379	39	40	41	$\Psi(h_m^2,h_f^2)$	106	89	78
		507	352	270	$\Psi(h_m^2)$	983	750	581
		364	376	386	$\Psi(h_f^2)$	283	293	303
h_f^2	0.428	49	50	52	$\Psi(h_m^2,h_f^2)$	116	99	88
		507	352	272	$\Psi(h_m^2)$	963	737	571
		370	383	394	$\Psi(h_f^2)$	286	296	306
	0.477	59	61	64	$\Psi(h_m^2,h_f^2)$	125	109	98
		509	354	274	$\Psi(h_m^2)$	944	723	562

From table 2 it can be concluded that a quadratic function is not sufficient to approximate the likelihood surface. Clearly the curvature for males depends on the values of both the male and female heritability, and similarly $\Psi(h_f^2)$ depends on both the heritability values. Fitting a cubic function in (h_f^2,h_m^2) to the complete 9 by 9 grid showed a fairly good approximation of the likelihood surface (not presented). The relative contributions from males and females at the IAM maximum can be derived from the (0.379,0.379) grid in table 2: $w_f = (383 + 41)/735 = 0.58$ and similarly $w_m = 0.42$. The sum of the elements of the curvature matrix (735) corresponds to the observed information in analysis I. The total observed information for model III is approximately 1000 (see tables 1 and 2).

We considered the 2nd differentials for a hierarchical balanced mating design with records on progeny only that gives a h_m^2 of 0.28 and a h_f^2 of 0.477. The second part of table 2 shows the curvature matrix for a design with 575 sires, 7 dams per sire and 2 progeny per dam. The structure is only a crude approximation of the structure in the data set, but a similar pattern is observed; the female heritability curvature is relatively constant, while the male heritability curvature depends heavily on the value of the male heritability.

DISCUSSION

One argument for a smaller h_m^2 than h_f^2 is because of selection in males. Alternatively, an increased variance might be expected as North-American sires are from a different population. Another explanation for the difference may be a Genotype by Environment (GxE) interaction: since most daughter-dam pairs have records in the same herd, a potential GxE effect would be contained in the daughter-dam covariance. Also there could be a cytoplasmic effect (Bell et al, 1985; Freeman, 1990), which causes a larger covariance between daughter and dams than between sire and offpring. The present data set was not suitable to investigate this effect, since the pedigrees could not be traced far enough back.

Various authors (e.g. Smith and Graser, 1986;Graser *et al*, 1987) have suggested a quadratic approximation of the likelihood surface to obtain asymptotic variances when the 2^{nd} differentials or the expectation thereof are not a by-product of the estimation algorithm. However, in data analysis and simulation it has been found that a quadratic approximation sometimes does not produce sensible results, in particular when many random effects are estimated (Meyer, 1989). A cubic approximation would produce better results, since the 2^{nd} differentials are still functions of the parameter values. However, for many random components, for example in a multiple trait situation, this would involve inverting a matrix of order $(1+p)(p(p+5)/6 + 1)$, where p is the number of random effects in the model. If for example p=5, the order of the matrix would be 56 and for p=10 it would be 286.

In the hierarchical case the heritability estimates are ratios of linear functions of mean squares. As the variances of mean squares depend on their expected value, it is no surprise that the curvature depends on the values of the parameters. One suggestion worth investigation would be to use transformations of the parameters to perhaps speed up convergence and ease interpretation. Candidate transformations are the z-transformation of Fisher (1921) and that of Wilson and Hilferty (1931).

The extended model has been suggested to validate existing models, but it is an open question whether the extended IAM will lead to faster genetic progress. It has applications in other areas, for example for testing if genetic variances are homogeneous in different parts of selection experiments.

ACKNOWLEDGEMENTS

P.V. acknowledges financial support from the Milk Marketing Boards in the U.K. We thank Professor W.G. Hill for helpful comments.

REFERENCES

BELL, B.R., MCDANIEL, B.T., and ROBISON, O.W. 1985. J. Dairy Sci. 68:2038-2051.
FISHER, R.A. 1921. Metron 1 part 4:1-32.
FREEMAN, A.E. 1990. Digest of the British Cattle Breeders Club 45:33-36. Cambridge, U.K.
GRASER, H-U, SMITH, S.P. and TIER, B. 1987. J. Anim. Sci. 64:1362-1370.
MEYER, K. 1989. Genet. Sel. Evol. 21:317-340.
PATTERSON, H.D. and THOMPSON, R. 1971. Biometrika 58:545-554.
SMITH, S.P. and GRASER, H-U. 1986. J. Dairy Sci. 69:1156-1165.
THOMPSON, R. 1977. Biometrics 33:497-504.
VISSCHER, P.M. 1990. Proc. 4th World Congr. Genet. Appl. Livest. Prod..
WIGGANS, G.R., MISZTAL, I. and VAN VLECK, L.D. 1988. J. Dairy Sci. 71 suppl. 2:54-69.
WILSON, E.B. and HILFERTY, M.M. 1931. Nat. Acad. Sci. 17:694-699.