Livestock Production Science, 28 (1991) 273–290
Elsevier Science Publishers B.V., Amsterdam

Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds

P.M. Visscher[a], R. Thompson[b] and W.G. Hill[a]
[a]*Department of Genetics, University of Edinburgh, West Mains Road, Edinburgh EH9 3JN, UK*
[b]*AFRC Institute of Animal Physiology and Genetics Research, Roslin, EH25 9PS, Midlothian, UK*

(Accepted 21 November 1990)

ABSTRACT

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.*, 78: 273–290.

Genetic and environmental variances for fat yield were estimated for 26 large Holstein Friesian pedigree herds in the UK, using an Animal Model (AM). Individual herd heritability estimates ranged from 0.03 to 0.80, but a single test against the combined herd estimate (0.379) was not significant ($P > 0.05$). Testing variance estimates for each herd at a constant heritability resulted in highly significant heterogeneous (phenotypic) variance estimates. A quadratic approximation of the likelihood surface at the maximum likelihood estimates seemed to over-estimate (log)likelihood differences when testing different parameter values. Three different transformations of the data to reduce the heterogeneity of variance were tried. Pre-adjustment for the within-herd or herd–year–season phenotypic variance was most effective in reducing heterogeneity of variance, and seems the most practical way to deal with heterogeneity of variance for an AM genetic evaluation.

Keywords: dairy cattle; animal model; heritabilities; heterogeneity of variance; maximum likelihood.

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 (Best Linear Unbiased Prediction; 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,

0301-6226/91/$03.50

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, for example, Hill et al., 1983; Lofgren et al., 1985; Brotherstone and Hill, 1986; Mirande and Van Vleck, 1985; Boldman and Freeman, 1988, 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. 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.

TABLE 1

Summary statistics for individual herd parameters

Parameter	Mean	Min	Max	Q1	Q3	STDEV
Mean	212.4	170.3	263.6	189.9	228.1	26.85
Raw σ_p^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			
$r(\sigma_p$, 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.

$r(\sigma_p$, mean): empirical correlation between herd means and herd phenotypic standard deviations.

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 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.

METHODS

The following linear model was fitted

$$\mathbf{y}=\mathbf{Xb}+\mathbf{Zu}+\mathbf{e} \text{ and}$$

$$v(\mathbf{y})=\mathbf{ZAZ}'\sigma_a^2+\mathbf{I}\sigma_e^2=\mathbf{ZGZ}'+\mathbf{R}$$

where **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 HF and lactation length were fitted as covariables. Three seasons of 4 months were defined as December–March, April–July and August–November, which correspond to the season definition for the current UK 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)

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

where $\mathbf{C}$ = full rank submatrix of the coefficient matrix (the matrix containing the left-hand side of the Mixed Model Equations (Henderson, 1973)) and $\mathbf{y}'\mathbf{P}\mathbf{y}$ = residual sum of squares, with $\mathbf{P}$ a projection matrix.

The estimations were carried out using a REML program written by Meyer (1989), which uses an iterative (grid) 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, for example, 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, for example, Mood et al., 1973), for which $2(L1-L2)$ is assumed to follow a Chi-squared 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, 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

$$y_{ij}^{c}=y_{ij}[\mathrm{sd_p}/\mathrm{sd_i}]$$

with $\mathrm{sd_p}$=population s.d., $\mathrm{sd_i}$=s.d. for HYS i and y^c=adjusted trait.

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, for example, 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.

RESULTS

The results from the individual herd analyses are presented in Table 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

TABLE 2

Individual herd REML estimates and LR tests

Herd	h^2	s.e.(h^2)	Δ_1	σ_a^2	s.e.(σ_a^2)	Δ_2	σ_e^2	s.e.(σ_e^2)	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	525.7	168.4	0.4	447.5	132.8	0.4	0.0
7	0.42	0.30	0.0	56.4	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)										
1–26	0.35	0.03	33.7	23.5$^{!}$	2.4	99.7$^{!}$*	444.8	22.2	58.4*	268.4*
Single combined herd analysis estimates, used as H_0 values										
1–26	0.379			324.5			532.3			

$\Delta_{1,2,3} = -2$(difference in log-likelihood) at ML estimate and H_0 value.
$\Delta_4 = -2$(difference log-likelihood) for variances at $h^2 = 0.379$.
*Significant at less than 5% level.
$^{!}$Extrapolation error; estimate is not significant.
$^{!}$*Extrapolation error; estimate is significant at 5% level.
The variance estimates are in units kg^2.

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. 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

Results of combined herd analyses

Analysis	σ_a^2	σ_e^2	h^2	s.e.(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

TABLE 4

Summary of likelihood ratio tests

		Analyses			
		I	II	III	IV
Separate LR test for each herd estimate					
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
$\sigma_p^2 \mid H^2$	$P<0.05$	13	1	8	9
	$P<0.01$	11	0	8	6
Single LR test					
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**
$\sigma_p^2 \mid H^2$		268.4**	16.9	151.7**	154.7**

Analyses: I=standard, II=date adjusted for within-HYS phenotypic standard deviation, III=log transformation, IV=square root transformation.
Values for separate herd LR tests are numbers of estimates which are significantly different from the H_0 values.
Values for single LR tests are -2[difference log-likelihood].
(!) LR values are overestimates because of extrapolation errors (see text).
$\sigma_p^2 \mid H^2$=ML estimate of the phenotypic variance at the heritability estimate from the combined herd analysis (Table 3), which is used as H_0 value.
**$P<0.01$.

A summary of the LR tests is given in Table 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) 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).

The likelihood differences in columns 4, 7 and 10 of Table 2 were from likelihood comparisons with the ML estimates from the combined herd analysis, which were 324.5, 532.3 and 0.379 for the genetic variance, environmental variance and the heritability, respectively (see Table 3). For two data sets the heritabilities and genetic variances were different from the overall estimate ($P<0.05$). In six cases the environmental variance was significantly different from 432.3. 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 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 ($P<0.01$) LR (205.7, for 25 degrees of freedom).

In general, the quadratic approximation over-estimated 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 two heritabilities, two genetic variances and two environmental variances (from four 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). 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 was different from the overall estimate. A single LR test indicated no significance for all three parameters ($P>0.05$).

For the log transformed data, two heritabilities, one genetic variance and three environmental variances for individual herds differed ($P<0.05$) from the H_0 values. However, assuming equal heritabilities (0.372), eight 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.

DISCUSSION

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 UK 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 (Hill et al., 1983; Mirande and Van Vleck, 1985; Brotherstone and Hill, 1986; Boldman and Freeman, 1988; Meinert et al., 1988), 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), 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 a 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) 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 UK 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 pre-corrected 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; Boldman and Freeman, 1988; Meinert et al., 1988), 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). Superior cows in low producing herds would therefore be over-evaluated 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. If 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.

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.

For illustration, following Visscher and Thompson (1990), 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 5. Clearly the curvature changes with different parameter values. Visually, this is demonstrated in Fig. 1, which represents likelihood contours for various combinations of the values of genetic and environmental variances from Table 5, using both exactt 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 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.

Various authors (e.g. Smith and Graser, 1986; Graser et al., 1987) have

TABLE 5

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

	σ_a^2					
σ_e^2	0.30	0.35	0.40	0.45	0.50	
0.50	1092.86	849.89	673.69	541.85	440.77	$\Psi(\sigma_a^2)$
	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	892.77	695.86	552.59	445.08	362.42	$\Psi(\sigma_a^2)$
	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	737.27	575.44	457.33	368.50	300.07	$\Psi(\sigma_a^2)$
	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	614.52	479.90	381.40	307.16	249.88	$\Psi(\sigma_a^2)$
	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	516.27	403.13	320.14	257.49	209.08	$\Psi(\sigma_a^2)$
	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: σ_a^2=40, σ_e^2=60.

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 approxima-

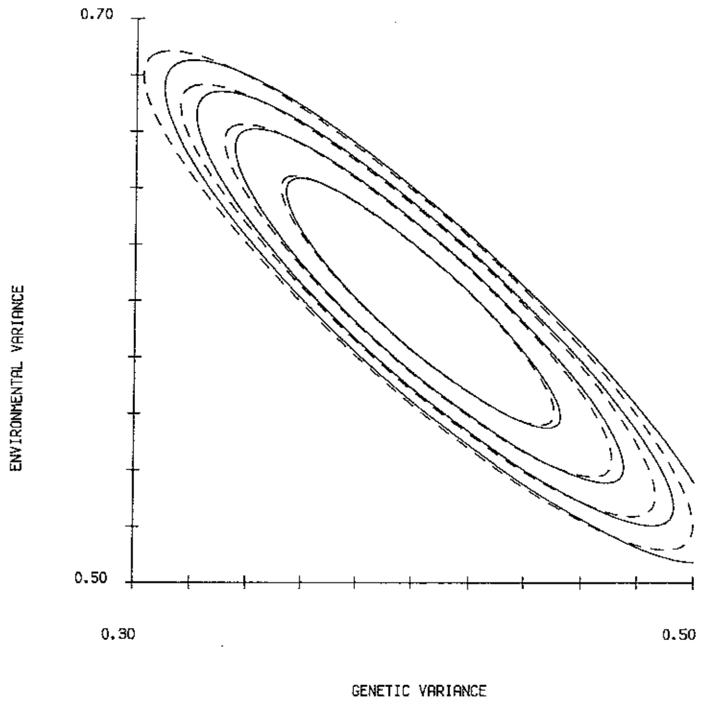

Fig. 1. Likelihood contours for balanced design. Parameters: $\sigma_e^2=0.60$; $\sigma_a^2=0.40$ ($h^2=0.40$; $\sigma_p^2=1.0$). Data: on 10 progeny of 100 sires. Difference between subsequent contour lines is 0.10 log-likelihood. Solid curves: contours for exact likelihoods. Dashed curves: contours for likelihoods obtained from a quadratic approximation of the likelihood surface around the ML values.

TABLE 6

Exact and predicted likelihood ratios (LR) for a balanced design

$h^2(t)$	H^2(ML)									
	0.20		0.30		0.40		0.50		0.60	
	LR1	LR2	LR1	LR2	LR1	LR2	LR1	LR2	LR1	LR2
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.
LR1 = exact LR from likelihood curve.
LR2 = predicted LR from quadratic around the maximum.

tion 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.

ACKNOWLEDGEMENT

P.V. acknowledges financial support from the Milk Marketing Boards in the UK.

REFERENCES

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.

Brotherstone, S. and Hill, W.G., 1986. Heterogeneity of variance amongst herd for milk production. Anim. Prod., 42: 297–303.

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.

Fisher, R.A., 1956. Statistical Methods and Scientific Inference. Oliver and Boyd, Edinburgh, London, 180 pp.

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

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.

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

Henderson, C.R., 1973. Sire evaluation and genetic trends. Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. J. Lush. ADSA and ASAS, Champaign, IL, pp. 10–41.

Hill, W.G., 1984. On selection among groups with heterogeneous variance. Anim. Prod., 39: 473–477.

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., 39: 59–68.

Interbull, 1986. Procedures for International Comparisons of Dairy Sires — Current Practice and Evaluation of Methods. Bulletin no. 1, Uppsala, Sweden, 28 pp.

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.

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., 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.

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, NY, 564 pp.

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.

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.

Searle, S.R., 1979. Notes on variance component estimation. Paper BU-673-M of Biometrics Unit, Cornell University, USA, 146 pp.

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.

Swalve, I. 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.

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., 73: 3053–3060.

Vinson, W.E., 1987. Potential bias in genetic evaluations from differences in variation within herds. J. Dairy Sci., 70: 2450—455.

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. Proc. 4th World Congr. Genet. Appl. Livest. Prod., 14: 233–236.

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.

RESUME

Visscher, P.M., Thompson, R. et Hill, W.G., 1991. Estimation des variances génétiques et de milieu pour la production de matières grasses dans des troupeaux particuliers et analyse de l'hétérogénéité de variance entre troupeaux. *Livest. Prod. Sci.*, 28: 273–290 (en anglais).

Les variances génétiques et non génétiques pour la production de matière grasse ont été estimées sur 26 grands troupeaux Holstein Friesian selon le modèle animal (MA). Les coefficients d'héritabilité ont varié de 0,03 à 0,80 suivant les troupeaux mais ils n'ont pas différé significativement ($P > 0,05$) d'une estimé globale (0,379). En supposant une héritabilité constante, les estimées de variances phénotypiques ont varié très significativement d'un troupeau à l'autre. L'approximation de la fonction de vraisemblance par une fonction quadratique a probablement contribué à rendre trop sensible le test d'hétérogénéité de variances. Trois transformations différentes sur les données de base ont été essayées en vue de réduire l'hétérogénéité de variance. Un préadjustement pour la variance phénotypique intra-troupeau ou intra-troupeau–année–saison a été le plus efficace pour réduire l'hétérogénéité de variance. Il paraît être le moyen le plus pratique de traiter l'hétérogénéité de variance en vue d'une indexation suivant le modèle animal.

KURZFASSUNG

Visscher, P.M., Thompson, R. und Hill, W.G., 1991. Schätzung von genetischer und Umweltvarianz der Fettmenge in einzelnen Herden und Untersuchungen zur Heterogenität der Varianzen zwischen Herden. *Livest. Prod. Sci.*, 28: 273–290 (auf englisch).

Die genetische Varianz und die Umweltvarianz der Fettmenge wurden für 26 große Holstein Fresian Herdbuchbetriebe in den UK mit einem Tiermodell geschätzt. Die Schätzwerte für die Heritabilität in den einzelnen Herden lag zwischen 0,03 und 0,80, aber einfache Tests gegenüber der Gesamtheritabilität (0,379) ergaben keine signifikanten Abweichungen ($P > 0{,}05$). Ein Test der Varianzschätzwerte jeder Herde bei konstanter Heritabilität ergab eine hoch signifikante Heterogenität der phänotypischen Varianzen. Wenn verschiedene Parameterwerte getestet werden, scheint eine quadratische Approximation der Wahrscheinlichkeitsoberfläche um das Maximum die Loglikelihood-Unterschiede zu überschätzen. Drei verschiedene Transformationen der Daten zur Reduzierung der Heterogenität der Varianzen wurden untersucht. Eine Vorkorrektur auf die phänotypische Varianz innerhalb der Herde oder der Herdenjahrsaison war am besten geeignet, diese Heterogenität zu verringern; sie erscheint der praktischste Weg zu sein, um eine Heterogenität der Varianzen in einem Tiermodell zu berücksichtigen.