Anim. Prod. 1992, **55:** 321-329

0003-3561/92/19590321$02·00

Heterogeneity of variance and dairy cattle breeding

P. M. Visscher† and W. G. Hill

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT

Abstract

The effects of heterogeneity among herds in phenotypic variance (σ^2_P) and in heritability (h^2) on the accuracy of evaluation of dairy sires and cows are discussed with reference to estimates from field data. It is argued that heterogeneity of σ^2_P can be accounted for by using a (Bayesian) regression procedure. While the same procedure could be used to correct for heterogeneity of h^2, in most practical situations the sampling errors of the h^2 estimates for individual herds are likely to be so large that regressed estimates of h^2 will differ little from each other, so constant values can be assumed.

Keywords: *dairy cattle, genetic variance, heritability, selection.*

Introduction

In dairy cattle breeding, best linear unbiased prediction (BLUP; Henderson, 1973) is the standard way of predicting breeding values for sires, and is becoming so for jointly evaluating sires and cows using an animal model (Wiggans, Misztal and Van Vleck, 1988; Ducrocq, Boichard, Bonaiti, Barbat and Briend, 1990; Jones and Goddard, 1990). For the properties of BLUP to hold, particularly in a population undergoing selection, the exact variances and covariances of all the observations should be known and should be used in the genetic evaluation. Fitting an incorrect model or covariance structure may reduce accuracy and therefore response to selection. For computational reasons, and for lack of accurate parameter estimates, models are usually fitted with as few parameters as possible. For example, national animal model (AM) evaluations in the USA, Australia and France use a single trait repeatability model to describe the data, and fit a 'management group' as the main fixed effect.

One assumption often made for BLUP genetic evaluation is that variances are homogeneous across levels of fixed and random effects. There is, however, abundant evidence of heterogeneity of variance across environments, in particular across herds (see. for example, Hill, Edwards, Ahmed and Thompson, 1983; Lofgren, Vinson, Pearson, and Powell, 1985; Mirande and Van Vleck, 1985; Brotherstone and Hill, 1986; Boldman and Freeman, 1990; Dong and Mao, 1990; Short, Blake, Quaas and Van Vleck, 1990). Failure to take account of this heterogeneity of variance may lead to inaccurate and biased predictions of breeding values, although the potential loss in response due to heterogeneity induced prediction error (HIPE) depends on the true (unknown) covariance structure of the data.

The aim of this study was to review some of the theory and results from parameter estimation concerning heterogeneity of variance in dairy cattle, and to present some new material, including a derivation of loss in response to sire selection and an investigation into a practical way suggested previously for dealing with heterogeneous (phenotypic) variances. Most of the discussion is on heterogeneity of variance across herds.

Theory

Definitions

Let Y_{ij} be a record on the jth animal in the ith herd, and $Y_{ij} = FX + A_{ij} + E_{ij}$

---

† Present address: Department of Food and Agriculture, Livestock Improvement Unit, 166-176 Wellington Parade, East Melbourne, Victoria 3002, Australia.

where: FX are defined environmental effects (including herd effects); A_{ij} = breeding value; E_{ij} = environmental deviation. These have variances: $v(A_{ij}) = \sigma^2_{Ai}$, $v(E_{ij}) = \sigma^2_{Ei}$, $v(A_{ij} + E_{ij}) = \sigma^2_{Ai} + \sigma^2_{Ei} = \sigma^2_{Pi}$ and are uncorrelated, and the heritability or regression of breeding value on phenotype in herd i is

$$h^2_i = \beta_{Aij\,Pij} = \sigma^2_{Ai}/\sigma^2_{Pi}.$$

Prediction of breeding values for comparisons within a herd using only data from that herd depends solely on its parameters. For example, the within herd prediction for mass selection is

$$\hat{A}_{ij} = h^2_i\,(Y_{ij} - \bar{Y}_i).$$

In any national programme, predictions have to be made across herds, so we define $A^\star_{ij}$ = the breeding value in a 'national herd', which can be predicted from data in one herd using $\beta_{A^\star_{ij}Y_{ij}}$.

For overall evaluation using BLUP, the relationships between animals, the fixed effects structure and the variances of the Y_{ij} and covariances with $A^\star_{ij}$ are required over all herds. Several assumptions can be made about heterogeneity of variances across herds. Some are listed in Table 1, most of which were also given by Garrick and Van Vleck (1987). For example, for scenario 1 it is assumed that both additive genetic and environmental variances are homogeneous, and for scenario 2 that the within herd genetic and environmental variances are the same functions of the herd mean, so that heritabilities are homogeneous across herds. The scenario involving most parameters, but still assuming a genetic correlation of unity between breeding values across herds, is 6, for which it is assumed that variances and heritabilities are heterogeneous across herds. Scenario 7+ here includes any with non-unit genetic correlations across herds.

Table 1 *Assumptions about heterogeneity of parameters for records nested within herds*†

Scenario	$v(A_{ij})$	$v(E_{ij})$	h^2_i	ρ‡
1. Homogeneity	σ^2_A	σ^2_E	h^2	1
2. Homogeneity after scale transformation§	$\sigma^2_A f(\mu_i)$	$\sigma^2_E f(\mu_i)$	h^2	1
3. Heterogenous variances, homogeneous h^2	σ^2_{Ai}	σ^2_{Ei}	h^2	1
4. Heterogeneous environmental variances and h^2	σ^2_A	σ^2_{Ei}	h^2_i	1
5. Homogeneous genetic variances after scale transformation,heterogeneous environmental variances and h^2§	$\sigma^2_A f(\mu_i)$	σ^2_{Ei}	h^2_i	1
6. Heterogeneity	σ^2_{Ai}	σ^2_{Ei}	h^2_i	1
7+. Various, with genotype × environment interaction				< 1

† For symbols, see text.
‡ Genetic correlation between performance in herd and nationally.
§ $f(\mu)$ is an arbitrary function of the mean μ.

Accuracy of selection

If heterogeneity of variance is not taken into account, an incorrect model is fitted to the data; but the practical question in genetic evaluation is whether there is a sufficient loss of accuracy of selection to matter.

Mass selection. Results shown here are based on (obscure) derivations of Hill (1984). Given a set of k groups each of size n, an expression for the accuracy, r, in the standard environment of phenotypic selection on deviations from group mean $\bar{P}_{i.}$, is:

$$r = \{\sum_i\sum_j[P_{ij} - \bar{P}_{i.}),\, A^\star_{ij}]\}/\{\sum_i\sum_j\ (P_{ij} - \bar{P}_{i.})^2 \sum_i\sum_j A^{\star 2}_{ij}\}^{1/2}.$$

Taking the expectation over animals within groups (E_j),

$$E_j(r) = \sum_i E\{\sum_j[(P_{ij} - \bar{P}_{i.}),\, A^\star_{ij}]\}/$$

$$[\sum_i E\{\sum_j(P_{ij} - \bar{P}_{i.})^2\} \sum_i E\{\sum_j A^{\star 2}_{ij}\}]^{1/2}$$

$$= \sum_i\ [\mathrm{cov}(P_{ij} - \bar{P}_{i.},\, A^\star_{ij})]/[\sum_i[\mathrm{v}(P_{ij} - \bar{P}_{i.})]\sum_i[\mathrm{v}(A^{\star 2}_{ij})]]^{1/2}$$

$$= (1 - 1/n)^{1/2}\sum_i \mathrm{cov}(A_i, A^\star_i)/[\sum_i(\sigma^2_{Pi})\sum_i(\sigma^{2\star}_{Ai})]^{1/2},$$

which becomes, assuming a genetic correlation of unity between performances across groups and the same variance $\sigma^{2\star}_{Ai}$ for all groups,

$$E_j(r) = (1 - 1/n)^{1/2}\sum_i \sigma_{Ai}/[k\sum_i \sigma^2_{Pi})]^{1/2}.$$

Now taking expectations across groups, which may be a fixed or a random set,

$$E(r) = (1 - 1/n)^{1/2}\,\mathrm{E}(\sigma_{Ai})/[\mathrm{E}(\sigma^2_{Pi})]^{1/2}$$

$$\approx (1 - 1/n)^{1/2}\,(h_0 + \mathrm{cov}(h_i, \sigma^2_{Pi})/\sigma_{P0})/[1 + \mathrm{CV}^2(\sigma_P)]^{1/2}$$

$$= r_0\,[1 + \rho_{h,\sigma_p}\mathrm{CV}(h)\mathrm{CV}(\sigma_P)]/[1 + \mathrm{CV}^2(\sigma_P)]^{1/2} \qquad (1)$$

with $r_0 = (1 - 1/n)^{1/2}\,h_0$, if variances are homogeneous. CV is the coefficient of variation, $\sigma_{P0} = E(\sigma_P)$, $h_0 = \mathrm{E}(h_i)$ and ρ_{h,σ_p} is the correlation between h_i and σ_{Pi}. Formula (1) corresponds to formula (2) of Hill (1984).

In Table 2 predicted reductions or increases in accuracies are presented for different values of ρ, CV(h) and CV(σ_{Pi}), using (1). Van Vleck (1987) and Vinson (1987) give examples of losses in response to selection using the results from Hill (1984).

A somewhat surprising result from (1) is that if $\rho_{h,\sigma_P}\mathrm{CV}(h) > \mathrm{CV}(\sigma_P)$, i.e. heritabilities and phenotypic variances are positively correlated and dispersion of heritabilities is large relative to that of phenotypic variances, the expected accuracy is larger than under homogeneity of variances. The change in accuracy is small, however, unless CV(σ_P) is very large.

Finally, accuracy of mass selection can be calculated when heterogeneity of variance is taken into account for prediction of breeding values. For group i, the regression of $A^\star$ on P_{ij} is $h_i\sigma_{A\star}/\sigma_{Pij}$, so that the accuracy of selection becomes,

$$r = (1 - 1/n)^{1/2}\,[E(h_i^2)]^{1/2},$$

and its expectation over all groups,

$$E(r) \approx r_0\,(1 + \tfrac{1}{2}\mathrm{CV}^2(h)).$$

Therefore, if all the variances are known so that correct weights can be given to records in different groups, accuracy is somewhat higher if heritabilities vary among groups.

Sire selection. Suppose each sire has n progeny in each of k herds, and that herds differ in intra-class correlation (ICC), t_i. Using simple selection index theory assuming the t_i (usually taken to be $h_i^2/4$) and σ_{Pi} are known without error so optimal weights are used, it can be shown that the squared accuracy of selecting sires (repeatability) is,

$$r^2 = nk/\{nk + k/\sum_i [t_i/(1 - t_i)]\} \qquad (2)$$

$$= nk/\{nk + H(\lambda_i)\},$$

where $H(\lambda_i)$ is the harmonic mean of $\lambda_i = (1 - t_i)/t_i$.

Letting $\lambda_0 = (1 - t_0)/t_0$, the expectation of r^2 is approximately $E\,(r^2) \approx nk/[nk + \lambda_0 - (1 - 1/kt_0)\mathrm{CV}^2(t)/(1 - t_0)]$.

If all $t_i = t_0$, (2) reduces to the well known repeatability of a predicted breeding value from a progeny test with nk daughters, $r^2 = nk/(nk + \lambda_0)$. Formula (2) indicates that dispersion of λ_i leads to an increase in accuracy, since the harmonic mean is less than the arithmetic mean if the t_i vary. This result is analogous to the simple case of response to mass selection using data on repeat records of an individual; if they are uncorrelated response is proportional to $(h_1^2 + h_2^2)^{1/2}$, so dispersion of heritabilities gives a higher response to selection than equal heritabilities with the same mean.

If heterogeneity of variances is ignored, and all weights given to progeny means in different herds are assumed to be equal to $(1 - t_0)/t_0^{1/2}$ (see Hill *et al.*, 1983, for the appropriate weights given to progeny means in different groups), then the squared accuracy, r^2, of selection with balanced data is:

$$r_1^2 = nk/\{nk + k[\Sigma(1 - t_i)]/[(\Sigma\, t_i^{1/2})^2]\}. \qquad (3)$$

The expectation of r_1^2 is, approximately, using Taylor's expansion

$$E[r_1^2] \approx nk/[nk + \lambda_0/\{1 - \tfrac{1}{4}\,\mathrm{CV}^2\,(t)\}].$$

Hence, as anticipated, the accuracy is reduced if the t_i differ but are assumed to be homogeneous. The expected reduction in accuracy is very small, however. If this analysis is extended to incorporate variation among herds in σ_{Pi}, yet equal weights are applied, there is a further loss in accuracy which is a function of CV(σ_{Pi}) and of the correlation between t_i and σ_{Pi} over herds; but the loss in accuracy remains small if many herds are involved.

Table 2 *Accuracy of mass selection with heterogeneity of variance relative to the accuracy with homogeneity of variance as a function of the coefficient of variation of phenotypic standard deviation (CV(σp)) and of root heritability (CV (h)) and the correlation between them* (ρ_{h,σ_p})

		CV(σ_P)				
		0·1	0·2	0·3	0·4	0·5
ρ_{h,σ_P}	CV(h)	Relative accuracy × 100				
−0·5	0·1	99·0	97·1	94·3	91·0	87·2
	0·2	98·5	96·1	92·9	89·1	85·0
	0·3	98·0	95·1	91·5	87·3	82·7
	0·4	97·5	94·1	90·0	85·4	80·5
	0·5	97·0	93·2	88·6	83·6	78·3
0·0	any	99·5	98·1	95·8	92·8	89·4
0·5	0·1	100·0	99·0	97·2	94·7	91·7
	0·2	100·5	100·0	98·7	96·6	93·9
	0·3	101·0	101·0	100·1	98·4	96·2
	0·4	101·5	102·0	101·5	100·3	98·4
	0·5	102·0	103·0	103·0	102·1	100·6

BLUP

For most genetic evaluations in practice using BLUP with a sire model or an animal model, designs are unbalanced. With an animal model, different types of relatives contribute to the prediction of breeding values so it is not obvious how to predict losses in accuracy if heterogeneity of variance is ignored.

Garrick and Van Vleck (1987) investigated losses in response due to heterogeneous variances between three environments corresponding to herd mean production groups by deterministic simulation using two sets of parameters. Responses were compared to the response using a multitrait model which was assumed to be the correct (true) model. For one parameter set phenotypic variances and heritabilities increased from the low to the high mean production group, and ignoring heterogeneity of variance led to a negligible reduction in asymptotic rate of response to selection. For a second parameter set, genetic (sire) variances were constant across environments and residual variances decreased with increased herd mean production levels, so that heritabilities decreased with increased phenotypic variances. This led to a reduction in the rate of response of approximately 3% when heterogeneous variances were ignored in the genetic evaluation. Garrick and Van Vleck thus concluded that, in practice, progeny testing schemes were robust to violations of assumptions regarding homogeneity of variances between environments.

Meuwissen and Van der Werf (1992) simulated data for a dairy population undergoing selection and investigated losses in response to selection when heterogeneous variances between herds were ignored in the genetic evaluation. Data were simulated either with or without heterogeneous variances and heritabilities were constant across herds. Breeding values were predicted using an animal model and response to selection was computed assuming there was homogeneity of variance. For both progeny testing and open nucleus schemes there were no significant losses in response when variances were heterogeneous rather than homogeneous. If a herd year effect was fitted in the model but not included in the simulation, there was a slight increase in the response to selection with the open nucleus scheme for the case of simulated heterogeneous variances; but it is not clear why this occurred. The number of potential bull dams was very large in their simulation, so reducing their number by selecting from a limited group of (high variance) herds may reduce intensity of selection only slightly. Furthermore, in the simulation of Meuwissen and Van der Werf (1992), (multiple) female progeny from elite matings had records in the same herd as their dams. Therefore after a few generations of selection the high variance herds are also the herds with the highest genetic mean, so that selection within this group of herds does not necessarily lead to a great loss in asymptotic rate of response to selection.

Field data analyses

Many results of analyses of heterogeneity of variance from field data have been reported using different populations of dairy cattle, different methods of variance component estimation, and different assumptions about heterogeneity of variance between herds (see Table 1). We refer only to results from analyses on Holstein-Friesian populations in the last decade. Unless stated otherwise, parameters were estimated using residual maximum likelihood (REML; Patterson and Thompson, 1971).

Grouping herds

Several studies have been reported in which herds were grouped according to their mean production or an estimate of their phenotypic variance, and parameters were estimated within and between groups. If herds are grouped according to their mean production this implies an assumption of scenarios 2 and 5 from Table 1. Results from those studies were discussed in a previous paper (Visscher, Thompson and Hill, 1991).

There have been relatively few analyses quantifying the genetic correlation of animal performances across herd groups (assuming scenario 7+). Hill *et al.* (1983) estimated genetic correlations of sire performances for milk production traits in herd groups split according to mean, phenotypic variance, or coefficient of variation (CV), using ANOVA type estimation methods. For all criteria of grouping herds and for all traits, genetic correlations did not differ significantly from unity. Dong and Mao (1990) estimated genetic correlations between sire performances in three herd groups (low, medium, high) split according to mean herd milk production level or herd standard deviation, for three separate 4-year time periods. All estimates were larger than 0·95, and for the most recent time period (1984 to 1987) estimates of genetic correlations across herd groups were essentially unity. Estimates of genetic correlations for milk yield, fat yield and fat content between pairs of states in the USA were reported by Carabaño, Wade and Van Vleck (1990). The lowest correlation was 0·93 (for fat yield between California and Wisconsin), and estimates for fat content exceeded 0·98. Although the average genetic correlation of performance between any two herds in the population was not investigated because herds were 'selected', it seems reasonable for practical purposes to assume that such a correlation is unity, thereby elimating scenario 7+ of Table 1.

Meyer (1987) estimated a sire × herd variance of approximately 0·03 of the total phenotypic variance, but no distinction could be made between a real genotype × environment interaction and an effect due to common treatment of half-sibs within herds. Garrick and Van Vleck (1987) found no reduction in response to selection when using 'correct' (simulated) sire and residual variances across three environments but assuming a genetic correlation of unity of performances across these environments, although simulated genetic correlations ranged from 0·86 to 0·97.

Using individual herd data

To investigate scenarios 2 to 6, genetic and environmental parameters should be estimated for individual herds, but few analyses have been reported which quantify heterogeneity of variances between herds in dairy cattle.

Brotherstone and Hill (1986) estimated variances and coefficients of variation (CV) of individual herd phenotypic standard deviations for milk, fat and protein yield in Holstein-Friesians. Variances of the parameter of interest (θ_i, for example σ_{Pi}) were estimated using:

$$\begin{aligned}&\text{variance between } \theta_i\\ &= \text{empirical variance between } \hat{\theta}_i\\ &\quad - \text{average sampling variance of } \hat{\theta}_i \qquad (4).\end{aligned}$$

The estimated CV of herd σ_{Pi} ranged from 0·155 to 0·172 for milk, fat and protein yield, for both first and later lactation records.

Van Vleck and Dong (1988) and Visscher *et al.* (1991) used REML with an animal model to estimate genetic and environmental variances for individual herds. From a likelihood ratio (LR) test, Visscher *et al.* (1991) concluded that, although individual herd heritabilities were not significantly different from each other, herd phenotypic variances were highly significantly different. It was subsequently shown, however, that the statistical power of such a significance test to detect differences in heritabilities between herds can be very low because the standard errors of the heritability estimates are large relative to those of phenotypic variance (Visscher, 1992). Standard errors of heritability estimates for individual herds in these studies were approximately 0·15 (Van Vleck and Dong, 1988) and 0·19 (Visscher *et al.*, 1991).

An alternative way to estimate $v(\theta_i)$, the variance of the parameter of interest in the population of herds, is to use the LR statistic from comparing the maximum likelihood from estimating a single heritability estimate using all data (ML_0), with 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 this function is assumed to be quadratic in θ, then

$$L_i = a_i + c_i(\theta_i - \hat{\theta}_i)^2,$$

where a_i and c_i are constants and $\hat{\theta}_i$ is 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

$$\hat{\theta}_i = \theta_i + \varepsilon_i;\ \hat{\theta}_0 = \theta_0 + \varepsilon_0;\ v(\varepsilon_i) = v(\hat{\theta}_i \mid \theta_i),$$

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

$$\begin{aligned}&E\,\{2[\Sigma ML_i - ML_0]\}\\ &= E\{\Sigma[(\varepsilon_i - \varepsilon_0)^2/v(\varepsilon_i)] + \Sigma[(\theta_i - \theta_0)^2/v(\varepsilon_i)]\}\\ &\approx d + v(\theta_i)[\Sigma\, 1/v(\varepsilon_i)], \qquad (5)\end{aligned}$$

where d = degrees of freedom (= $k - 1$). Letting t = $2(\Sigma ML_i - ML_0)$, an estimate of $v(\theta_i)$ is therefore

$$\hat{v}(\theta_i) = (t - d)/[\Sigma\, 1/v(\varepsilon_i)] \qquad (6).$$

The empirical variance among the 26 heritability estimates obtained by Visscher *et al.* (1991) and their average sampling variance were found to be 0·035 and 0·039 respectively, giving a difference (equation 4), which was negative. These values confirm the outcome of a more elaborate likelihood ratio test in which heritability estimates were found not to differ significantly (Visscher *et al.*, 1991).

Equation (6) was checked for robustness against violations of assumptions using Monte Carlo simulation for a nested balanced half-sib design of sires within herds (for details see Visscher, 1992). Equation (6) was found to be reasonably accurate in estimating the variance of the parameters. For example, for a design of 10 herds, 100 sires per herd and 10 progeny per sire, the standard deviation of heritabilities between herds was simulated to be 0·010, and estimated (using equation 6) as 0·009. Similarly, for a design of 25 herds, 25 sires nested within herds, and 20 progeny per sire, the standard deviation and its estimate were 0·030 and 0·024, respectively. The downward bias in both cases seems to come from using the estimate of the heritability in each herd to approximate its ANOVA sampling variance: when the mean of the true heritabilities over all herds, or an estimate thereof, was used to calculate sampling variances of heritability estimates for individual herds, the estimates of the standard deviations for each design were 0·010 and 0·031 respectively. Using the likelihood ratio statistic of Visscher *et al.* (1991), the standard deviation of heritabilities was estimated using (6) and found to be 0·098, corresponding to a CV of 0·25. The relatively large difference between the estimates of the variances using (4) and (6) may be due to the fact that in heritability estimates from each herd are weighted equally in (4), whereas they are weighted according to their approximate sampling variances in (6).

Scale transformations

If scenarios 2 and 5 from Table 1 apply, a scale transformation will remove some or all of the heterogeneity of variance. For scenario 2, correcting data from different herds for their within-herd phenotypic standard deviation would also yield homogeneous genetic variances across herds since heritabilities are constant. Some practical aspects of this kind of standardization, as proposed by Hill (1984), are discussed in the next section. If the true correlation between herd mean and herd standard deviation is unity, a logarithmic transformation would also remove heterogeneity of variance. Published estimates of this correlation for milk yield traits are, however, in the range 0·4 to 0·5, departing from 1·0 further than can be accounted for by sampling (Mirande and Van Vleck, 1985; Boldman and Freeman, 1990; Visscher *et al.*, 1991), and implying that a log transformation would reverse the correlation to give a negative association between herd mean and herd standard deviation. A consequence of a log transformation would then be that more animals (e.g. bull dams) would be selected from the herds of low means. Whether this is justified depends on the heritabilities within these herds. For example, genetic variances are assumed to be homogeneous after a scale transformation in scenario 5, while phenotypic variances and heritabilities remain heterogeneous. Such a scenario was proposed by Garrick and Van Vleck (1987) and Boldman and Freeman (1990). If heritabilities for untransformed traits are higher in high mean groups, as found in several studies (e.g. Hill *et al.*, 1983; Boldman and Freeman, 1990; Dong and Mao, 1990), a log transformation would reduce the efficiency of selection. As pointed out by others (Garrick and Van Vleck, 1987; Boldman and Freeman, 1990), one should therefore be careful in applying a log transformation for genetic evaluation purposes in dairy cattle.

Bayesian regression

To account for sampling errors of parameter estimates, Hill (1984) and Brotherstone and Hill (1986) proposed that individual herd parameters be regressed to an overall *a priori* estimate, a standard Bayesian procedure in which the regression coefficient depends on the sampling variances of individual herd estimates and the variance of the parameters, assumed to be known *a priori*. This method has been used for cow index evaluation in Britain (G. J. T. Swanson, personal communication). The regression may be written:

$$\hat{\theta}_i^{\star} = \hat{\theta}_0 + \beta_i(\hat{\theta}_i - \hat{\theta}_0) \tag{7}$$

where $\hat{\theta}_i$ is the parameter estimate and $\hat{\theta}_i^{\star}$ its regressed estimate for herd i, and $\hat{\theta}_0$ is the overall (prior) estimate. The regression coefficient is $\beta_i = 1/(1 + \gamma_i)$, with $\gamma_i = \mathrm{v}(\hat{\theta}_i \mid \theta_i)/\mathrm{v}(\theta_i)$ being the ratio of the sampling variance to the variance of the parameter, or less formally, the ratio of variance 'within' and 'between' parameters θ. Thus if $\hat{\theta}_i$ is estimated accurately from a large number of observations, $\mathrm{v}(\hat{\theta}_i \mid \hat{\theta}_i)$ and γ_i are small, so the weighted estimate $\hat{\theta}_i^{\star}$ is close to $\hat{\theta}_i$; whereas if $\hat{\theta}_i$ has a high sampling error, the weighted estimate is close to the population value, θ.

Let us consider the effect of the regression as proposed by Brotherstone and Hill (1986) if there are true differences between individual herd heritabilities, but sampling variances are too large to detect them. One suggestion is to assume a value for the coefficient of variation of heritabilities among individual herds and apply (7). Using the mean estimate of the 26 heritability estimates as an estimate of the population value (the 'prior' estimate), i.e. of the mean of all true individual herd heritabilities, equation (7) was applied to the 26 heritability estimates from Visscher *et al.* (1991). Results are presented in Table 3. For a CV among (true) heritability values of 0·10 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 (4), it is unlikely to be worthwhile taking individual herd heritability estimates for many herds because the regressed values are nearly homogeneous. Even for the estimate of the CV of 0·25 (last row Table 3), the standard deviation of regressed heritabilities using data from large pedigree herds was only 0·05. Smaller herds would give heritability estimates with such large standard errors that the regressed values

Table 3 *Regressions of individual herd heritability estimates to an overall mean, assuming different coefficients of variation for the unobserved heritabilities*†

CV	$\bar{\gamma}$	$\bar{\beta}$	$\bar{\theta}^{\star}$	s.d.($\theta^{\star}$)	Range($\theta^{\star}$)
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

† Using estimates fromVisscher *et al.* (1991): γ and β are variance ratio and regression, from (7); $\theta^{\star}$ is h^2 after regression; s.d. ($\theta^{\star}$) and range ($\theta^{\star}$) are s.d. and range of regressed h^2; $\mathrm{v}(\theta) = (\mathrm{CV})^2(\hat{\theta}_0)^2$, with $\hat{\theta}_0 = \bar{\theta} = 0{\cdot}388$.

would be very close to the overall mean (or prior estimate).

The two methods to estimate $v(\theta_i)$ were applied to the phenotypic variances for fat yield estimated by Visscher *et al.* (1991), and were found to give similar results. Using equations (4) and (7), the estimate of CV was 0·248 (from a mean of 835 kg^2), the standard deviation of regressed variances was 189 kg^2, and the average regression coefficient was 0·86. Using equation (6), the corresponding values were 0·228, 190 kg^2 and 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 (6) was applied to the phenotypic standard deviations. The estimates of their mean and CV were 28·6 kg and 0·124 respectively, similar to estimates for fat yield reported by Brotherstone and Hill (1986). After regressing individual herd phenotypic standard deviations (average β was 0·86), the s.d. of the regressed values was 3·4 kg. Comparing the estimates of the CV for phenotypic variances and heritabilities again shows that the finding from Visscher *et al.* (1991) of homogeneous heritabilities and heterogeneous phenotypic variances can follow solely from relatively high sampling variances, and conveys little about true differences between heritabilities.

Whether it is worthwhile to estimate phenotypic variances for each herd in the population regularly and apply formula (7) to regress the estimates to some prior mean depends on the relative magnitude of the true variance and the sampling variance of the parameters. Consider a balanced design with d degrees of freedom to estimate the phenotypic variance for each of k herds. Assuming normality, the sampling variance of an estimate of the phenotypic variance is

$$v(\hat{\theta}^2_{Pi}) = 2\sigma^4_{Pi}/d \text{ or } v(\hat{\theta}_i) = 2\theta_i^2/d,$$

and, for use in (7),

$$\beta_i = 1/[1 + 2\theta_i^2/dv(\theta)].$$

Hence, assuming $E[\theta_i^2] \approx \theta_0^2$,

$$\bar{\beta} = E[\beta_i] \approx \{CV\theta_i^2)\}/[\{CV\theta_i^2)\} + 2/d] \qquad (8)$$

Assuming $\hat{\theta}_0 \approx (\Sigma\hat{\theta}_i)/k$, the mean CV of *regressed* estimates is, for large k,

$$E[CV\hat{\theta}_i^{\star})] \approx \bar{\beta}\{CV(\theta^2) + (2/d)\}^{1/2} \qquad (9)$$

For various values of d and CV ($\sigma_{P_i}{}^2$), the CV of the regressed estimates are presented in Table 4, using (8) and (9). Clearly, for small d and CV, heterogeneity of variance could be ignored.

Gianola, Foulley, Fernando, Henderson and Weigel (1992) give a formal empirical Bayesian method for estimation of heterogeneous variance. Essentially, they proposed that individual herd (genetic and environmental) variances be estimated by

$$\hat{\sigma}^{2\star} = \frac{d\hat{\sigma}^2 + d_0\sigma_0{}^2}{d + d_0} = \sigma_0^2 + \frac{d}{d + d_0}(\hat{\sigma}^2 - \sigma_0^2),$$

with d the degrees of freedom for the initial estimate of the herd variance. The parameter d_0 was termed 'degrees of believe' of the average population value by Gianola *et al.*, and is inversely proportional to the prior variance of variances. This method has been used to standardize individual herd phenotypic

Table 4 *Coefficients of variation for regressed individual herd phenotypic variance estimates for a balanced design*†

	CV(θ)									
	0·1		0·2		0·3		0·4		0·5	
d	$\bar{\beta}$	CV(θ*)	$\bar{\beta}$	CV(θ*)	$\bar{\beta}$	CV(θ*)	$\bar{\beta}$	CV(θ*)	$\bar{\beta}$	CV(θ*)
10	0·05	0·02	0·17	0·08	0·31	0·17	0·44	0·27	0·56	0·37
20	0·09	0·03	0·29	0·11	0·47	0·21	0·62	0·31	0·71	0·42
30	0·13	0·04	0·38	0·12	0·57	0·23	0·71	0·34	0·79	0·44
40	0·17	0·04	0·44	0·13	0·64	0·24	0·76	0·35	0·83	0·46
50	0·20	0·04	0·50	0·14	0·69	0·25	0·80	0·36	0·86	0·46

† d is d.f. $\bar{\beta}$ is the average regression coefficient, from (8). CV(θ) is CV of true phenotypic variances. CV(θ*) is CV of regressed estimates.

variances for the Australian animal model evaluation since 1984 (L. Jones, personal communication), assuming $d_0 = 20$. Hence the value of d_0 can either be estimated from data, for example using equation (4), or just assumed. Equating the two regression coefficients, β_i and $d/(d + d_0)$, gives

$$d_0 = 2/\mathrm{CV}^2_{(\theta)}, \text{ or } \mathrm{CV}_{(\theta)} = (2/d_0)^{1/2}.$$

Discussion

All cases described have dealt with heterogeneity of variance across levels of fixed effects, particularly herds. In dairy cattle there is evidence, however, of different heritabilities estimated from different kinds of relatives (see Visscher and Thompson, 1992, and references therein). Estimates of heritabilities from paternal half-sibs are usually lower than from daughter-dam regression or using an animal model and a statistical model to deal with such heterogeneity of (genetic) variance has been given (Visscher and Thompson, 1992).

Theoretical results have been derived for mass selection and sire selection using a progeny test. In only the former case, i.e. for bull-dam selection, does heterogeneity of variance substantially reduce accuracies. There are other situations, however, in which the heterogeneity can raise problems when comparisons are made between animals tested in different places.

Nucleus (MOET, multiple ovulation and embryo transfer) herds are now being established in which bulls are evaluated from records of their dam or sibs in one MOET herd (Nicholas and Smith, 1983). Since the (commercial) aim of MOET herds is to produce (sell) animals of high genetic merit for the overall population, it is important that the estimated breeding values of animals from these herds predict the performance of progeny in the 'national' herd, which requires that the genetic mean of the herd and the variance about it are both well estimated. There is a distinction between the dairy MOET herd, in which bulls from it are likely to be compared directly with, for example, progeny tested sires from the national population, and the herds or flocks of pig and poultry breeding companies from which an overall product rather than individual animals assessed relative to some national standard are marketed.

It may be expected that the management in MOET herds is such that the mean production level (and therefore herd variance) and heritability are above average. With no correction for herd phenotypic variance for a national genetic evaluation the predicted breeding values of high MOET bulls would therefore be likely to be too high whereas if the correction were solely for phenotypic variance, they would be likely to be too low.

In this study the criterion for assessing the efficiency of selection was the correlation between predicted and true breeding values. Another criterion is the variance of predicted breeding values. If this variance is too high for a group of animals, say bulls, then the best bulls of that group are not as good as suggested by their predicted breeding values. This is particularly important if bulls are compared across different schemes or groups, for example local and foreign bulls, or progeny tested and MOET bulls. Hence an incorrect variance of predicted breeding values for one scheme means that too many or too few bulls are selected from that scheme, so that the overall accuracy of selection is decreased.

It is concluded that, for practical purposes, individual herd phenotypic variances should be estimated and regressed, by a factor (see equation 7) depending on the number of records per herd, to a prior estimate. It seems unlikely, however, that heritability estimates for individual herds will be available for use in heterogeneity adjustment, and it is therefore suggested that herd heritabilities are assumed to be homogeneous. If, on average, herd heritabilities are associated with their mean production or their phenotypic variance, individual herd heritabilities can be estimated using their mean or phenotypic variance. A similar regression as proposed before can be applied, taking into account the magnitude of the correlation between heritability and mean (or variance), and hence the sampling error. The present adjustment for heterogeneity of variance for the US genetic evaluation uses a relationship between herd heritability and herd phenotypic variance estimated from data before 1984 (Wiggans and VanRaden, 1991). Other schemes could be adopted incorporating this information.

Acknowledgements

This work was supported by a grant from the Milk Marketing Board. We thank M. Goddard for many helpful comments.

References

Boldman, K. G. and Freeman, A. E. 1990. Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. *Journal of Dairy Science* **73:** 503-512.

Brotherstone, S. and Hill, W. G. 1986. Heterogeneity of variance amongst herds for milk production. *Animal Production* **42:** 297-303.

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. *Journal of Dairy Science* **73:** 173-180.

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

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. *Journal of Dairy Science* **73:** 1945-1955.

Garrick, D. J. and Van Vleck, L. D. 1987. Aspects of selection for performance in several environments with heterogeneous variances. *Journal of Animal Science* **65:** 409-421.

Gianola, D., Foulley, J. L., Fernando, R. L., Henderson, C. R. and Weigel, K. A. 1992. Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *Journal of Dairy Science* In press.

Henderson, C. R., 1973. Sire evaluation and genetic trends. *Proceedings of the animal breeding and genetics symposium in honor of Dr. J. Lush,* pp. 10-41. American Society of Animal Science, American Dairy Science Association, American Poultry Science Association, Champaign, Illinois.

Hill, W. G. 1984. On selection among groups with heterogeneous variance. *Animal Production* **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. *Animal Production* **36:** 59-68.

Jones, L. P., and Goddard, M. E. 1990. Five years experience with the animal model for dairy cattle evaluations in Australia. *Proceedings of the fourth world congress on genetics applied to livestock production, Edinburgh, Vol. XIII,* pp. 382-385.

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. *Journal of Dairy Science* **68:** 2737-2739.

Meuwissen, T. H. E. and Van der Werf, J. H. J. 1992. Impact of heterogeneous within herd variances on dairy cattle breeding schemes. *Livestock Production Science* In press.

Meyer, K. 1987. Estimates of variances due to sire × herd interactions and environmental covariances between paternal half-sibs for first lactation dairy production. *Livestock Production Science* **17:** 95-115.

Mirande, S. L. and Van Vleck, L. D. 1985. Trends in genetic and phenotypic variances for milk production. *Journal of Dairy Science* **68:** 2278-2286.

Nicholas, F. W. and Smith, C. 1983. Increased rates of genetic changes in dairy cattle by embryo transfer and splitting. *Animal Production* **36:** 341-353.

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

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 yields of grade Holstein cows. *Journal of Dairy Science* **73:** 3312-3320.

Van Vleck, L. D. 1987. Selection when traits have different genetic and phenotypic variances in different environments. *Journal of Dairy Science* **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. *Journal of Dairy Science* **71:** 3040-3046.

Vinson, W. E. 1987. Potential bias in genetic evaluations from differences in variation within herds. *Journal of Dairy Science* **70:** 2450-2455.

Visscher, P. M. 1992. Power of likelihood ratio tests for heterogeneity of intraclass correlation and variance in balanced half-sib designs. *Journal of Dairy Science* **75:** 1320-1330.

Visscher, P. M. and Thompson, R. 1992. Comparisons between genetic variances estimated from different types of relatives in dairy cattle. *Animal Production* **55:** 315-320.

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. *Livestock Production Science* **28:** 273-290.

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

Wiggans, G. R. and VanRaden, P. M. 1991. Method and effect of adjustment for heterogeneous variance. *Journal of Dairy Science* **74:** 4350-4357.

(*Received* 17 *December* 1991—*Accepted* 11 *June* 1992)