Original article

Univariate and multivariate parameter estimates for milk production traits using an animal model. II. Efficiency of selection when using simplified covariance structures

PM Visscher [1,*], WG Hill [1], R Thompson [2]

[1] *University of Edinburgh, Institute of Cell, Animal and Population Biology,*
West Mains Road, Edinburgh EH9 3JT;
[2] *AFRC Institute of Animal Physiology and Genetics Research,*
Roslin EH25 9PS, Midlothian, UK

Summary – 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. Previously estimated (co)variances were assumed to be the true population parameters. A modified repeatability model, *ie* assuming a genetic correlation of unity between performances across lactations but allowing for different heritabilities and variances in different lactations, was used to investigate the efficiency of selecting sire on their progeny performance. For most combinations of heritabilities in first and later lactations, selecting sires using the repeatability model was nearly 100% efficient in terms of accuracy of selection compared to a general multivariate model. The predicted response to selection using a modified repeatability model was approximately 10% too high. It was found that applying a transformation to make new traits in lactation 1 uncorrelated at the phenotypic and genetic level to milk, fat and protein yield in later lactations, and assuming that 3 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.

dairy cattle / milk production / selection index / genetic parameters / repeatability / canonical transformation

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

Résumé – Utilisation du modèle animal pour l'estimation des paramètres univariates et multivariates concernant les caractères de production laitière. II. Efficacité de la sélection après utilisation d'une structure simplifiée de covariance. *L'effet d'une simplification de la structure des covariances pour la production de lait, de matière grasse et de matière protéique a été analysé en utilisant la théorie des index de sélection. Les estimées antérieures de (co)variances ont été supposées être les vrais paramètres de la population. Un modèle "répétabilité" modifié, c'est-à-dire supposant une corrélation génétique de 1 entre les performances des différentes lactations mais permettant des héritabilités et des variances différentes à chaque lactation, a été utilisé pour analyser l'efficacité de la sélection des taureaux à partir de la performance de leur descendance. Pour la plupart des combinaisons d'héritabilité en première lactation ou lactations ultérieures, la sélection des taureaux en utilisant le modèle répétabilité a permis à peu près la même précision de la sélection que le modèle général multivariate. La réponse prédite à la sélection selon le modèle répétabilité modifié a été surévaluée d'environ 10%. On a appliqué aux lactations ultérieures une transformation rendant indépendants en première lactation, tant au niveau phénotypique que génétique, les caractères de production laitière, de matière grasse, de matière protéique et on a supposé que les 3 nouvelles variables étaient toujours indépendantes. Une telle procédure a été hautement efficace en termes de précision de la sélection comparée à la précision d'un modèle général multivariate. Cette transformation est recommandée pour une évaluation nationale selon le BLUP, puisqu'elle peut mieux tenir compte de la sélection que des analyses séparées des productions de lait, matière grasse et matière protéique.*

bovin laitier / production laitière / index de sélection / paramètre génétique répétabilité / transformation canonique

INTRODUCTION

In a previous paper (Visscher and Thompson, 1992), 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 yet feasible. In practice, therefore, simplified 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, b; Ducrocq *et al*, 1990; Jones and Goddard, 1990).

Selection index theory is used to investigate the loss in accuracy of selection when simplified covariance structures are used to predict breeding values. A second aim is to investigate how to reduce the dimensionality of the above MV prediction problem to a manageable size without a great loss in accuracy.

MATERIAL

As reported previously (Visscher and Thompson, 1992), the 9×9 genetic covariance matrix of M, F and P in lactations 1–3 was found to be not positive 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 I. An alternative way to summarise the covariance matrices is to calculate eigenvalues and eigenvectors of the matrix $\mathbf{P}^{-1}\mathbf{G}$. Eigenvectors then represent linear combinations of the original traits which are independent of each other at the genetic and phenotypic level (Hayes and Hill, 1980; Meyer, 1985). The corresponding eigenvalues of the newly created traits, called canonical variates (*eg* Anderson, 1958; ch 12), have been termed canonical heritabilities (Hayes and Hill, 1980; Meyer, 1985). Table II shows the eigenvalues and eigenvectors of matrix $\mathbf{P}^{-1}\mathbf{G}$, where $\mathbf{P}$ and $\mathbf{G}$ are the 9×9 phenotypic and genetic covariance matrices for milk, fat and protein yield in lactations 1–3 ($M_1F_1P_1M_2F_2P_2M_3F_3P_3$), calculated from parameters in table I. 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:

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

Table I. Scaled and rounded parameter estimates for milk, fat and protein yield in lactation 1–3[1].

	M_1	F_1	P_1	M_2	F_2	P_2	M_3	F_3	P_3
M_1	39	75	91	86	58	78	84	55	72
F_1	89	36	82	61	86	70	60	84	66
P_1	97	90	36	78	65	86	78	63	82
M_2	42	36	39	33	62	88	95	64	83
F_2	35	41	37	89	32	75	67	94	76
P_2	42	40	44	97	91	31	88	77	96
M_3	38	32	35	44	36	42	30	59	86
F_3	30	35	31	36	41	38	89	29	73
P_3	36	34	37	43	39	46	97	91	31
V_p^2	1.0	1.0	1.0	1.42	1.45	1.48	1.51	1.59	1.58

[1] Heritabilities ($\times 100$) on diagonals, genetic correlations ($\times 100$) above and environmental correlations ($\times 100$) below diagonals. [2] Phenotypic variance.

Therefore, one canonical trait, $\{-0.04M_1 + 0.15F_1 \therefore + 0.09P_3\}$, was found to have a heritability of -0.03. The negative eigenvalue resulted mainly from the contrast of individual yield traits in lactations 1, 2 and 3 ((M_2–M_3) + (F_1+F_2–F_3) + (P_2–P_1)). 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

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

Eigenvalue	M_1	F_1	P_1	M_2	F_2	P_2	M_3	F_3	P_3
				Corresponding Eigenvector					
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

row of table II). This was expected, given the very high genetic correlations for yield traits in lactations 2 and 3 (see table I).

ANALYSIS

Selection index theory

Let:
$\mathbf{u} = q \times 1$ vector of breeding values of q traits;
$\mathbf{a} = q \times 1$ vector of (marginal) economic values for q traits;
$\mathbf{H} = \mathbf{u}'\mathbf{a} =$ aggregate breeding value;
$\mathbf{x} = p \times 1$ vector of sources of information on an individual
(for example phenotypic observations, daughter averages, predicted breeding values);
$\mathbf{b} = p \times 1$ vector of index weights;
$\mathbf{I} = \mathbf{b}'\mathbf{x} =$ index value used to predict $\mathbf{H}$;
$\mathbf{P} = v(\mathbf{x}); \mathbf{G} = \mathrm{cov}(\mathbf{x}'\mathbf{u})$ and the symbol $\wedge$ added to a scalar or matrix indicates an estimate thereof.

For index calculations standard results were used (see, for example, Sales and Hill, 1976).

Equations for the responses are using $\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}\mathbf{a}$ for the optimal index, and $\widehat{\mathbf{b}} = \widehat{\mathbf{P}}^{-1}\widehat{\mathbf{G}}\mathbf{a}$ for an index using estimates of P and G :

$$\mathbf{R} = (\mathbf{a}'\mathbf{G}'\mathbf{P}^{-1}\mathbf{G}\mathbf{a})^{1/2} \tag{1}$$

$$\widehat{\mathbf{R}} = (\mathbf{a}'\widehat{\mathbf{G}}'\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{G}}\mathbf{a})^{1/2} \tag{2}$$

$$\mathbf{R}^* = (\mathbf{a}'\widehat{\mathbf{G}}'\widehat{\mathbf{P}}^{-1}\mathbf{G}\mathbf{a})(\mathbf{a}'\widehat{\mathbf{G}}'\widehat{\mathbf{P}}^{-1}\mathbf{P}\widehat{\mathbf{P}}^{-1}\widehat{\mathbf{G}}\mathbf{a})^{-1/2} \tag{3}$$

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

If new traits are created which are linear combinations of the observations, $\mathbf{y} = \mathbf{W}'\mathbf{x}$, *ie* variables in vector $\mathbf{y}$ are a linear combination of the variables in vector $\mathbf{x}$, then

$$\mathbf{R} = [\mathbf{a}'\mathbf{G}'\mathbf{W}\,(\mathbf{W}'\mathbf{P}\mathbf{W})^{-1}\mathbf{W}'\mathbf{G}\mathbf{a}]^{1/2} \qquad [4]$$

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.

$\alpha_i = \varepsilon_i \sigma_i$, 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, $\varepsilon_i = (0.8)^{i-1}$, 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 (from table I). If it is further assumed that the covariance of any observation with the breeding value in lactation 3 is equal to the covariance of that observation with breeding values in later lactations, *ie* the corresponding rows of matrix $\mathbf{G}$ 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 $4.0 (\Sigma(1.25)(0.8)^i = 4.0$, for $i = 2,3\ldots)$. Similarly for the case where only 2 lactations are considered and assuming second and later lactations, breeding values have equal covariances with observed phenotypes, $\mathbf{a}' = [1.0\ 5.0]$. Economic values for traits within a lactation were varied to reflect different breeding goals.

For all subsequent calculations it is assumed that observations in later lactations are included in the genetic evaluation only if observations on previous lactations are available.

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, since the genetic correlation between performances across lactations (r_g) is less than one, and indirectly through a better data structure (better "connectedness"). Assuming $\mathbf{a}' = [1\ 1\ 4]$ for either milk, fat or protein yield in lactations 1–3, and using the relevant parameters for any of these traits from table I, 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 was 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 modified repeatability model (*eg* assuming $r_g = 1$) is assumed for breeding value prediction instead of the "true" MV covariance structure. This was investigated for 3 selection indices:

$I_1 =$ phenotypic index, *ie* sources of information are phenotypic observations on individuals;

$I_2 =$ sire index; sources of information are daughter averages of sires in different lactations;

$I_3 =$ 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, *eg* 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 considered, 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 were different amounts of information on the 2 cohorts,

I_1:

Cohort 1: phenotypic observation in lactation 1

Cohort 2: phenotypic observations in lactations 1 and 2

I_2:

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 $(n_2 < n_1)$

I_3:

Cohort 1: sire index based on n first lactation progeny, dam index based on sire index of dam and dam's first lactation record, cow'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, cow's first and second lactation records.

Parameters used for the example with 2 lactations and 2 cohorts were (from table I): $\mathbf{a}' = [1\ 5]$; $r_g = 0.85$; $r_p = 0.55$; 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 (from table I) 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 this repeatability model differs from that usually used because heritabilities and phenotypic variances are not necessarily equal in different lactations.

Responses to selection were calculated using equations (1), (2) and (3). Given any set of parameters the optimal proportion of animals to be s 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 III.

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 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 expected response. This may be seen as a very simple illustration that one should be cautious when using predicted 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 3 indices used, subsequent calculations were only performed for the case of mass selection.

Table III. Relative responses to selection ($\times 100$) from using incorrect parameters in selection indices[1,2,3,4,5,6].

Index			I_1			I_2			I_3	
$\widehat{h}_1^2$	$\widehat{h}_2^2$	α_1	$\mathbf{R}^*/\mathbf{R}$	$\mathbf{R}^*/\widehat{\mathbf{R}}$	α_1	$\mathbf{R}^*/\mathbf{R}$	$\mathbf{R}^*/\widehat{\mathbf{R}}$	α_1	$\mathbf{R}^*/\mathbf{R}$	$\mathbf{R}^*/\widehat{\mathbf{R}}$
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

[1] $\widehat{h}_1^2$ and $\widehat{h}_2^2$ are heritabilities for first and second lactation performance used in selection index calculations. [2] $\mathbf{R}^*/\mathbf{R}$ = achieved response as percent of the maximum response. [3] $\mathbf{R}^*/\widehat{\mathbf{R}}$ = achieved response as percent of the predicted response. [4] α_1 = proportion of animals selected from age group 1. [5] Indices (see also text): I_1 = index for mass selection
I_2 = sire index
I_3 = cow index.[6]Economic values:$a' = [1.0\ 5.0]$.

Multiple trait multiple lactation considerations

Suppose the breeding goal is a linear combination of 9 traits (M_1, F_1, P_1, M_2, F_2, P_2, M_3, F_3, P_3), which is thought to be a good indicator of lifetime economic production since third and later lactation performances are assumed to be highly correlated. Then, choosing a set of economic values and using parameters from table I, the relative accuracy of selection for different indices which use different amounts of information can be investigated. For 3 different sets of economic values these relative accuracies were calculated, and results are presented in table IV. The economic values for M_1, F_1 and P_1 in the second breeding goal (H_2) are similar

to first lactation economic weightings used in practical selection indices in Europe (*eg* Wilmink, 1988). H_3 reflects a more "progressive" breeding goal with selection only on protein production. Results from table IV show that approximately 20% accuracy is lost when only one observation is used to predict the aggregate breeding value. Results for H_2 and H_3 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 (H_3). For all 3 breeding goals, the expected gain in accuracy for sire selection when adding progeny means on multiple traits is expected to be smaller than for mass selection.

Table IV. Accuracies of selection indices for mass selection as proportion ($\times 100$) of the accuracy using observations of M, F and P in lactations 1–3[1].

Traits fitted in selection index:	Breeding goal		
	H_1	H_2	H_3
M_1	84.4	57.0	73.3
F_1	79.1	71.2	63.9
P_1	83.4	69.6	78.8
M_2	80.2	57.6	74.1
F_2	74.4	72.7	64.3
P_2	81.8	74.1	82.0
M_3	76.8	55.0	72.9
F_3	70.1	71.8	60.2
P_3	80.0	75.9	83.9
M_1, F_1, P_1	85.9	78.4	79.4
M_2, F_2, P_2	82.8	83.9	83.2
M_3, F_3, P_3	80.3	89.2	87.8
M_1, M_2, M_3	95.6	66.9	86.9
F_1, F_2, F_3	89.8	86.1	75.3
P_1, P_2, P_3	97.0	86.9	96.7
M_1, F_1, P_1, M_2, F_2, P_2	95.6	90.8	91.5

[1] Breeding goals:
For H_1, $\mathbf{a}' = [1\ 1\ 1\quad 1\ 1\ 1\quad 1\ 1\ 1]$
$\quad\quad H_2$, $\mathbf{a}' = [-1.1\ 1.0\ 2.2\quad -1.1\ 1.0\ 2.2\quad -4.4\ 4.0\ 8.8]$
$\quad\quad H_3$, $\mathbf{a}' = [0\ 0\ 1\quad 0\ 0\ 1\quad 0\ 0\ 4]$

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

Proportionality considerations

One possible way 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 explains 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 (see for example Lin and Smith, 1990). In particular, it would be convenient if one linear transformation could be found that reduced the prediction problem of 9 highly correlated traits (milk, fat and protein yield in lactations 1–3) to that of 3 independent new traits, since in that case 3 separate analyses would contain all the information and would (therefore) account for selection. One approach is to investigate if submatrices of the 9×9 covariance matrix are proportional to each other, since that would indicate that a suitable transformation may exist to reduce the dimensionality of the prediction problem. With observation on p traits in l lactations, this suggests testing if the covariance matrix of lp traits, $\mathbf{V}$, may be written as $\mathbf{V} = \mathbf{K} \otimes \mathbf{V_h}$, where $\mathbf{V_h}$ is a submatrix (or a transformation thereof) describing a (co)variance block of p traits within or between lactations, $\mathbf{K}$ is a $l \times l$ symmetric matrix of proportionality or scaling constants, and $\otimes$ is the direct product operator (Searle, 1966). 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 (or scaling) matrix multiplied by a transformation matrix. In the appendix a general framework is presented to test for proportionality of covariance matrices using likelihood ratio tests and to estimate constants of proportionality, using multivariate normal theory applied to observed and expectations of moment matrices.

Unfortunately, the data from table I were found unsuitable for a likelihood ratio test. Obviously the additive genetic covariance matrix ($\mathbf{A}$) and the environmental covariance matrix ($\mathbf{E}$) from table I are not independent moment matrices; $\mathbf{A}$ and $\mathbf{E}$ are highly correlated and the determinant of $\mathbf{A}$ is zero. An alternative is to transform $\mathbf{A}$ and $\mathbf{E}$ into a between and within sire covariance matrix ($\mathbf{B}$ and $\mathbf{W}$), 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 $\mathbf{E}$ and $\mathbf{A}$ 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 pursued.

Using parameter estimates from table I, however, some inference with respect to proportionality may be drawn. One (obvious) choice for the transformation matrix $\mathbf{V_h}$ is a canonical transformation on M_1, F_1, and P_1. This transformation was calculated (see Hayes and Hill, 1980; and Meyer, 1985, for computations) and the matrix was used to transform the traits within second and third lactations. Let:

$$\mathbf{V_{g11}} = \mathbf{v_g} \begin{bmatrix} \mathbf{M_1} \\ F_1 \\ P_1 \end{bmatrix}, \text{ and } \mathbf{V_{p11}} = \mathbf{v_p} \begin{bmatrix} \mathbf{M_1} \\ F_1 \\ P_1 \end{bmatrix},$$

the 3×3 genetic ($\mathbf{V_{g11}}$) and phenotypic ($\mathbf{V_{p11}}$) covariance matrices of milk, fat and protein yield in lactation 1. Then the transformation matrix of milk, fat and protein yield in lactation one, $\mathbf{Q_1}$, was chosen such that:

$$\mathbf{Q_1 V_{g11} Q_1'} = \mathbf{D} \text{ and } \mathbf{Q_1 V_{p11} Q_1'} = \mathbf{I}$$

with elements of diagonal matrix $\mathbf{D}$ equal to eigenvalues of the matrix ($\mathbf{V_{p11}^{-1} V_{g11}}$).

Using $\mathbf{Q_1}$, the vector of observations,

$$\mathbf{y'} = [M_1F_1P_1 \quad M_2F_2P_2 \quad M_3F_3P_3] = [\mathbf{y_1' y_2' y_3'}],$$

was transformed using:

$$\mathbf{y_c} = \mathbf{Qy} = \begin{bmatrix} \mathbf{Q_1} & 0 & 0 \\ 0 & \mathbf{Q_1} & 0 \\ 0 & 0 & \mathbf{Q_1} \end{bmatrix} \begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \mathbf{y_3} \end{bmatrix}$$

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 9×9 correlation matrices and the heritabilities of the 9 new traits ($\mathbf{y_c}$) are shown in table V. Off-diagonals in all 3×3 blocks were small, indicating that one transformation matrix created 3 nearly independent variates with for each new 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 [A7]. This assumed the observed covariance matrices $\mathbf{E}$ and $\mathbf{A}$ were moment matrices, but degrees of freedom did not need to be specified. For $\mathbf{A}$ and $\mathbf{E}$, the estimates of proportionality matrix $\mathbf{K}$, $\mathbf{K_a}$ and $\mathbf{K_e}$ respectively, were:

$$\widehat{\mathbf{K}}_\mathbf{a} = \begin{bmatrix} 1.00 & 1.03 & 1.06 \\ & 1.55 & 1.32 \\ symm & & 1.74 \end{bmatrix}, \quad \widehat{\mathbf{K}}_\mathbf{e} = \begin{bmatrix} 1.00 & 0.60 & 0.58 \\ & 1.52 & 0.85 \\ symm & & 1.69 \end{bmatrix}$$

For example, the estimate of the average scaling factor for genetic variances for M, F and P on the transformed scale (transformation such that new variates are uncorrelated) in lactation 3 was 1.74, and the estimate of the average scaling factor for environmental covariances between the transformed variates in lactations 1 and 3 was 0.58. If proportionality is assumed, the 9×9 MV prediction problem may be reduced to 3 independent 3×3 multivariate predictions or to 3 independent evaluations with a repeatability model. Using the breeding goals defined previously the efficiency of the reduction in dimensionality was calculated for mass selection, conditional on the parameters in table I being the true population parameters. Thus the parameters from table V were used for index calculations with all off-diagonals of all 3×3 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 VI. The relative accuracy when using the first 3 canonical variates

Table V. Parameters on transformed scale after applying the canonical transformation matrix from lactation 1 to lactations 2 and 3[1,2].

	c_{11}	c_{12}	c_{13}	c_{21}	c_{22}	c_{23}	c_{31}	c_{32}	c_{33}
c_{11}	64	0	0	79	5	−3	78	3	−8
c_{12}	0	48	0	11	85	−3	10	82	1
c_{13}	0	0	35	−5	8	85	−8	8	83
c_{21}	64	2	−11	66	9	−15	77	4	−16
c_{22}	4	38	11	0	52	8	4	71	10
c_{23}	−3	0	43	−3	3	30	−13	11	98
c_{31}	64	3	−1	71	6	−5	65	4	−24
c_{32}	3	34	4	5	43	5	3	54	17
c_{33}	−1	−2	36	−7	7	44	1	−6	29

[1] Heritabilities ($\times 100$) on diagonals, genetic correlations ($\times 100$) above and environmental correlations ($\times 100$) below diagonals. [2] c_{ij} = Transformed variate j in lactation i.

is slightly lower than the corresponding accuracy using the original first 3 variates (M_1, F_1 and P_1) from table III because the genetic covariance structure between the canonical traits in lactation one and transformed variates in later lactations was simplified (off-diagonals of 3×3 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 3 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.

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 new traits for each lactation which are the sum of the phenotypic observations weighted by the corresponding economic values in the aggregate breeding value. Using the notation from equation [4], $y = \mathbf{a}'\mathbf{x}$, where variables in $\mathbf{x}$ are, for example, observations for M_1, F_1 and P_1. Relative accuracies were calculated using equation [4], fitting first lactation yield traits, first and second lactation yield traits, and all yield traits in vector $\mathbf{x}$, respectively. Results are presented in table VII.

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 $y_1 = \mathbf{w}'_1 \mathbf{x}_1$, for $\mathbf{x}'_1 = [M_1 F_1 P_1]$, and $y_2 = \mathbf{w}'_2 \mathbf{x}_2$, for $\mathbf{x}'_2 = [M_2 F_2 P_2]$, then in the selection index framework this would be fitting $\mathbf{y} = \mathbf{W}'\mathbf{x}$ as used for equation [4]. Using $\mathbf{x}_1$ and $\mathbf{x}_2$ as above, and $\mathbf{x}'_3 = [M_3 F_3 P_3]$, 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 VII.

Finally, 3 separate selection indices could be calculated from using ($M_1 M_2 M_3$), ($F_1 F_2 F_3$), and ($P_1 P_2 P_3$) respectively, and these 3 indices could be combined into one predicted aggregate breeding value. This approach is analogous to the

Table VI. Accuracies of selection indices for mass selection as proportion ($\times 100$) of the accuracy using observations of M, F and P in lactations 1–3, assuming proportionality of covariances blocks between traits across lactations[1,2].

		Breeding goal	
	H_1	H_2	H_3
Multivariate model using traits:			
c_{11}, c_{12}, c_{13}	85.8	78.2	79.1
c_{21}, c_{22}, c_{23}	82.0	83.0	82.3
c_{23}, c_{22}, c_{23}	78.5	87.1	85.4
All transformed variates	98.5	97.8	97.6
Repeatability model using traits:			
c_{11}, c_{12}, c_{13}	85.7	78.3	79.1
c_{21}, c_{22}, c_{23}	81.6	83.5	82.2
c_{31}, c_{32}, c_{33}	78.1	87.2	85.5
All transformed variates	98.2	96.4	96.7

[1] Breeding goals:

For H_1, $\mathbf{a}' = [1\ 1\ 1\quad 1\ 1\ 1\quad 1\ 1\ 1]$

$\quad H_2$, $\mathbf{a}' = [-1.1\ 1.0\ 2.2\quad -1.1\ 1.0\ 2.2\quad -4.4\ 4.0\ 8.8]$

$\quad H_3$, $\mathbf{a}' = [0\ 0\ 1\quad 0\ 0\ 1\quad 0\ 0\ 4]$

[2] c_{ij} = Transformed variate j in lactation i.

way selection indices for dairy cattle are usually calculated: breeding values are calculated for milk, fat and protein yield separately, and the 3 estimated breeding values are subsequently combined in a selection index (*eg* Dommerholt and Wilmink, 1986; Wilmink, 1988). Separate selection indices for milk yield, fat yield and protein yield were calculated using either the "true" (multivariate) covariance structure or the modified repeatability model (r_g across lactations set to unity), and the accuracy of using the 3 indices to predict the aggregate breeding value was calculated. For the construction of each of the 3 indices it was assumed that the traits in the index were identical to the traits in the aggregate breeding values. For example, the index using M_1, M_2 and M_3 was calculated as $\mathbf{P_m^{-1}G_m a_m}$, where $\mathbf{P_m}$ and $\mathbf{G_m}$ are the 3×3 phenotypic and genetic covariance matrices for milk yield in lactations 1–3, and $\mathbf{a_m}$ is a vector of economic values for these traits. The relative accuracies for this approach are shown in table VII.

Comparing results from tables IV, VI and VII 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 M_1, F_1 and P_1 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 H_2 and H_3 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 VII).

The last 2 rows of table VII indicate what accuracy was lost by ignoring within lactation information between milk, fat and protein yield for each lactation

Table VII. Accuracies of selection indices for mass selection as proportion ($\times 100$) of the accuracy using observations of M, F and P in lactations 1–3, when selecting on linear combinations of the observations (scalar y or vector $\mathbf{y}$)[1].

	Breeding goal		
	H_1	H_2	H_3
Section on single trait			
$y = \mathbf{w}'\mathbf{x}$			
$\mathbf{w} = \mathbf{a}^2$, using traits			
M_1, F_1, P_1	85.7	76.5	78.8
$M_1, F_1, P_1, M_2, F_2, P_2$	95.0	88.9	90.7
$M_1, F_1, P_1, M_2, F_2, P_2, M_3, F_3, P_3$	98.1	92.0	92.8
Selection on linear combination of observations ($\mathbf{y}$) within lactations			
$\mathbf{y} = \mathbf{W}'\mathbf{x}$			
$y_i = \mathbf{a}_i'\mathbf{x}_i^3$, 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
Selection on linear combination of observations ($\mathbf{y}$) across lactations			
$y = \mathbf{W}'\mathbf{x}$			
$y_i = \mathbf{b}_i'\mathbf{z}_i^4$, using traits			
y_1, y_2, y_3			
$\mathbf{b}_i$ from multivariate analysis	99.0	97.3	96.7
$\mathbf{b}_i$ from repeatability model[5]	99.0	96.6	96.1

[1] Breeding goals as in tables IV and VI; [2] $\mathbf{w} = \mathbf{a}$: weights are economic values; [3] $\mathbf{x}_i = [M_i \ F_i \ P_i]$ (for $i = 1, 2, 3$): milk, fat and protein yield for lactation i; [4] $\mathbf{z}_1' = [M_1 \ M_2 \ M_3]$; $\mathbf{z}_2' = [F_1 \ F_2 \ F_3]$; $\mathbf{z}_3' = [P_1 \ P_2 \ P_3]$, and $\mathbf{b}_i$ is calculated assuming the traits in index ($\mathbf{z}_i$) are the traits in $\mathbf{H}$; [5] Assuming r_g of traits across lactations is unity.

separately. In effect only average correlations between the traits within lactations are taken into account when combining separate indices for milk yield, fat yield and protein yield. The loss in accuracy was small and very similar to losses when using a repeatability model on the canonical variates (see table VI).

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. 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.

Use of 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 (second) derivatives. Still, calculations would then involve a 90×90 (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 9×9 $\mathbf{P}^{-1}\mathbf{G}$ matrix have similar biological explanations to the canonical variates from lactation 1, but now including comparisons between lactations (see table II).

Given that parameter estimates from table I 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. The loss in accuracy is very similar to the loss when information between traits within each lactation separately is ignored, as is common practice in dairy cattle selection indices. For practical (national) animal model BLUP evaluations, the proposed transformation is easy to implement.

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, this method of creating new traits from the available observations may be less efficient.

APPENDIX

Testing for proportionality of covariance matrices

Notation:

$\mathbf{M}$ = 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

$\mathbf{V}$ = $E(\mathbf{M})$; unknown PD covariance matrix of lp traits

$\mathbf{K}$ = symmetric matrix of proportionality constants of order l

$\otimes$ = direct product operator (see eg Searle, 1966), tr = trace operator

L = natural logarithm of likelihood.

Using standard multivariate theory (eg Anderson, 1958; ch 10),

$$L(\mathbf{M}, \mathbf{V}) = -1/2df[\log|\mathbf{V}| + \text{tr}(\mathbf{M}\mathbf{V}^{-1})] = -1/2df[\Sigma\log\lambda_i + \Sigma_{\gamma i}] \qquad [A1]$$

with λ_i = eigenvalue of $\mathbf{V}$, γ_i = eigenvalue of $\mathbf{M}\mathbf{V}^{-1}$.

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

$$\text{ML}(\mathbf{M}, \mathbf{V}) = -1/2df[\log|\mathbf{M}| + lp] \qquad [A2]$$

Suppose a moment matrix $\mathbf{M_0}$ is observed, and the null hypothesis is,

$\mathbf{H_0} : \mathbf{V_0} = E(\mathbf{M_0}) = \mathbf{kV}$ with $\mathbf{V}$ specified and $\mathbf{k}$ a constant.

Then,

$$L(\mathbf{M_0}, \mathbf{V_0}) = -1/2 \, df[(lp)\log(\mathbf{k}) + \log|\mathbf{V}| + (\Sigma_{\gamma_i})/\mathbf{k}] \qquad [A3]$$

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

$$\text{ML}_0(\mathbf{M_0}, \mathbf{V_0}) = -1/2 \, df[lp\log(\Sigma\gamma_i) - lp\log(lp) + \log|\mathbf{V}| + lp] \qquad [A4]$$

For the trivial case of $\mathbf{M_0} = \mathbf{kM}$ 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 ($= \mathbf{k}$). The likelihood ratio (LR) test statistic, $t = 2(\text{ML} - \text{ML}_0)$, asymptotically has a χ^2 distribution with degrees of freedom $[1/2lp(lp + 1) - 1]$. 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. $\mathbf{V_0}$ 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,

$$L_0(\mathbf{V_0}, \mathbf{M}) = -1/2df[\log|\mathbf{K} \otimes \mathbf{V_h}| + \text{tr}(\mathbf{M}(\mathbf{K} \otimes \mathbf{V_h^{-1}}))]$$
$$= -1/2df[l\log|\mathbf{K}| + p\log|\mathbf{V_h}| + \text{tr}(\mathbf{K}^{-1} \otimes \mathbf{I_p})(\mathbf{I_l} \otimes \mathbf{T_p^{-1'}})\mathbf{M}(\mathbf{I_l} \otimes \mathbf{T_p^{-1'}}))] \qquad [A5]$$

the trace in [A5] may be written as:

$$\text{tr}(\mathbf{P}(\mathbf{K}^{-1} \otimes \mathbf{I_p})\mathbf{P}'\mathbf{P}(\mathbf{I_l} \otimes \mathbf{T_p^{-1}})\mathbf{M}(\mathbf{I_l} \otimes \mathbf{T_p^{-1'}})\mathbf{P}')$$
$$= \text{tr}(\mathbf{I_p} \otimes \mathbf{K}^{-1})\mathbf{P}(\mathbf{I_l} \otimes \mathbf{T}^{-1})\mathbf{M}(\mathbf{I_l} \otimes \mathbf{T_p^{-1'}})\mathbf{P}')$$
$$= \text{tr}((\mathbf{I_p} \otimes \mathbf{K}^{-1})\mathbf{M}^*)$$
$$= \Sigma\text{tr}(\mathbf{K}^{-1}\mathbf{M_{ii}^*})$$

for $\mathbf{P}$ a permutation matrix and $\mathbf{M_{ii}^*}$ a $l \times l$ diagonal block of $\mathbf{M}^*$. Thus [A5] becomes:

$$L(\mathbf{V_0}, \mathbf{M}) = -1/2 df[l \log|\mathbf{K}| + p \log|\mathbf{V_h}| + \Sigma\text{tr}(\mathbf{K}^{-1}\mathbf{M_{ii}^*})] \qquad \text{[A6]}$$

Using [A6], the ML estimate of $\mathbf{K}$ is:

$$\widehat{\mathbf{K}} = (\Sigma\mathbf{M_{ii}^*})/l \qquad \text{[A7]}$$

ACKNOWLEDGMENT

REFERENCES

Anderson TW (1958) *An Introduction to Multivariate Statistical Analysis.* John Wiley and Sons, NY, 374 pp

Dommerholt J, Wilmink JBM (1986) Optimal selection response under varying milk prices and margins for milk prices and margins for milk production. *Livest Prod Sci* 14, 109-121

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

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

Hayes JF, Hill WG (1980) A reparameterisation of a genetic index to locate its sampling properties. *Biometrics* 36, 237-248

Jones LP, Goddard ME (1990) Five years experience with the animal model for dairy cattle evaluations in Australia. *In: Proc 4th World Congr Genet Appl Livest Prod, Edinburgh, Scotland July 23-27*, 13, 382-385

Lin CY, Smith SP (1990) Transformation of multitrait to unitrait mixed model analysis of data with multiple random effects. *J Dairy Sci* 73, 2494-2502

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

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

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

Searle SR (1966) *Matrix Algebra for the Biological Sciences*. John Wiley and Sons, NY

Visscher PM, Thompson R (1992) Univariate and multivariate parameter estimates for milk production traits using an animal model. I: Description and results of REML analyses. *Genet Sel Evol*, 24, 415-430

Wiggans GR, Misztal I, Van Vleck LD (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 GR, Misztal I, Van Vleck LD (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 JBM (1988) Selection on fat and protein to maximise profit in dairy herds. *Livest Prod Sci* 20, 299-316