REDUCED RANK ESTIMATES OF THE GENETIC COVARIANCE MATRIX FOR LIVE ULTRA-SOUND SCAN TRAITS

Karin Meyer

Animal Genetics and Breeding Unit[1], University of New England, Armidale, NSW 2351

SUMMARY

Multivariate restricted maximum likelihood analyses for a large data set comprising eight traits were carried out, estimating the leading 3, 4, 5 and 6 genetic principal components only. Traits were eye muscle area, percentage intra-muscular fat, and fat depth at the 12/13th rib and P8 sites, treating records on bulls and heifers or steers as different traits. The resulting, reduced rank estimates of genetic covariance matrices for analyses fitting 5 or 6 principal components agreed closely with an estimate from pooled, bivariate analysis. It is shown that reduced rank estimation can result in substantial reduction in computational requirements, compared to standard analyses fitting unstructured covariance matrices, and thus facilitate higher-dimensional multivariate analyses.
Keywords : Genetic parameters, beef cattle, scan traits, reduced rank, principal components

INTRODUCTION

In estimating genetic covariance matrices, analyses are very often limited to a few traits. Estimates of higher dimensional matrices are usually obtained by combining estimates from several, lower dimensional analyses. Limits on the dimensions of multivariate analyses are imposed by computational requirements. Whilst the number of effects fitted increases linearly with the number of traits considered, the number of non-zero elements in the mixed model equations, and thus of calculations required per likelihood evaluation in a restricted maximum likelihood (REML) analysis, increases much more rapidly. By and large, covariance matrices are assumed to be unstructured, so that for q traits there are $q(q+1)/2$ distinct elements of the matrix to be estimated. Maximising the associated log likelihood ($\log \mathscr{L}$) tends to become more difficult as the number of parameters increases. Moreover, sampling errors increase. Unless the traits considered are essentially uncorrelated, the corresponding covariance matrices have a number of eigenvalues close to zero. This implies that there are linear combinations of traits which contribute very little information and can be omitted (Kirkpatrick and Meyer 2004). Estimating the first m principal components (PCs) only reduces the number of parameters to $m(2q-m+1)/2$, and gives estimates of covariance matrices which have reduced rank m. This paper presents a multivariate analysis of eight traits, obtaining reduced rank estimates of the genetic covariance matrix by considering the leading genetic PCs only.

MATERIAL AND METHODS

Data. Data consisted of records for traits measured by live ultrasound scanning for Angus cattle in 36 herds which had 1000 or more animals with scan records. Traits considered were eye muscle area (EMA), fat depth at the 12th/13th rib (RIB), P8 fat depth (P8), and percentage intra-muscular fat (IMF), recorded from 300 to 700 days of age, with a single record per trait. Records on heifer or steers (.H) and bulls (.B) were treated as separate traits, yielding a total of 8 traits. After basic edits, the data comprised

Table 1. Characteristics of the data structure

Trait[A]	Heifers/Steers				Bulls			
	P8 (mm)	RIB (mm)	EMA (cm^2)	IMF (% $\times$10)	P8 (mm)	RIB (mm)	EMA (cm^2)	IMF (% $\times$10)
No. records	38,601	38,251	38,604	23,918	34,966	34,923	35,167	18,432
Mean	6.692	4.998	58.99	44.00	4.152	3.203	77.00	28.38
S.D.[B]	3.482	2.490	9.422	18.91	2.000	1.367	12.67	16.60
Age	506.7	506.2	507.3	524.8	489.1	488.5	489.2	496.1
No. CG[C]	1296	1295	1293	712	1568	1567	1569	819

[A] see text for abbreviations, [B] standard deviation, [C] contemporary group subclasses

$262,862$ records on $74,268$ animals, $34,649$ heifers, $35,345$ bulls and $4,274$ steers. Generally, all four measures for an animal were taken at the same time. However, IMF recording was introduced some time after the other traits. Hence, only 51.1% of bulls and 60.4% of heifers and steers had all four traits recorded, with most of the remainder having records for P8, RIB and EMA. In addition, there was a small number of animals with other combinations of traits, due to missing observations or deletion of dubious records. Further details are given in Table 1.

Analyses. Estimates of covariance components were obtained by REML. The model of analysis fitted contemporary groups (CG), birth type (single *vs.* twin) and a dam age class (heifer *vs.* cow) as fixed effects. CG were defined as herd-sex-management group-date of recording subclasses, with a further subdivison ("age slicing") if the range of ages in a subclass exceeded 60 days. In addition, age at recording for each sex and age of dam were fitted as linear and quadratic covariables. The only random effects fitted were additive genetic effects. Including pedigree information for animals with records and their parents up to four generations backwards resulted in a total of $103,467$ animals in the analysis.

Estimates of covariances matrices were obtained from eight-variate analyses, estimating the first 3, 4, 5 and 6 genetic principal components only, as described by Meyer and Kirkpatrick (2005). The residual covariance matrix was assumed to have full rank throughout, with the 16 residual covariances between traits measured on heifers or steers and bulls assumed to be zero, resulting in only 20 covariances to be estimated. In addition, 28 corresponding bivariate analyses were carried out to estimate correlations between all pairs of traits. Results were pooled using 'iterative summing of expanded part matrices' (Mäntysaari 1999), as implemented by Henshall and Meyer (2002).

RESULTS AND DISCUSSION

Characteristics of the reduced rank analyses are summarised in Table 2. In essence, fitting the first *m* PCs only reduced computational requirements of the eight-variate analysis to those of a corresponding $m-$variate analysis. The full rank analysis would have comprised 56 parameters to be estimated, and $837,915$ effects and 28.59×10^6 elements in the mixed model matrix. Factorisation would have created "fill-in" to yield a total of 186.9×10^6 non-zero off-diagonal elements, and would have required 212.6×10^9 operations. For our relatively large data set, $\log \mathscr{L}$ increased significantly for increasing numbers of PCs fitted. With a stringent penalty on the number of parameters, the corresponding Bayesian Information Criterion (BIC) too indicated that fitting six PCs was 'best'.
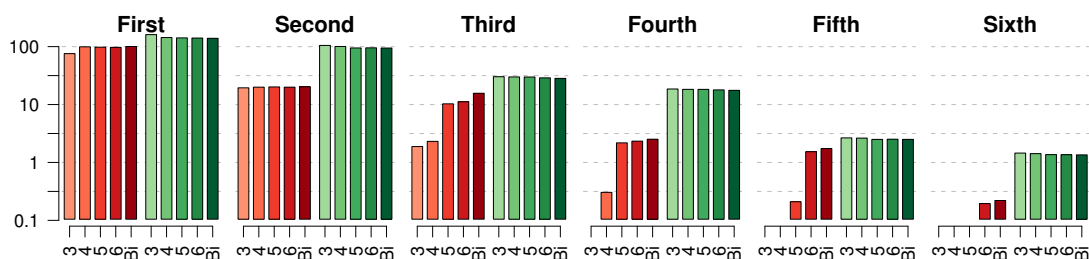
Figure 1. Estimates of the first six genetic (left) and residual (right) eigenvalues.

Estimates of the first 6 eigenvalues from different analyses are shown in Figure 1. The remaining eigenvalues from pooled bivariate analyses were 0.195 and 0.011 for genetic, and 0.494 and 0.221 for residual covariances. Overall, some repartitioning of genetic into residual variances was evident for reduced rank analyses. Whilst the first two eigenvalues differed comparatively little between analyses if at least 4 PCs were considered, estimates of third and fourth eigenvalues were substantially lower for analysis fitting 3 or 4 PCs only than for analyses considering more PCs. Zero residual covariances between traits measured on differents sexes resulted in estimates of the first two residual eigenvalues of similar magnitude, each being the first eigenvalue of an independent sub-block of the residual covariance matrix. Enforcing such structure in estimation may have affected repartitioning between genetic and residual components, and thus may have caused some abrupt changes in estimates of genetic eigenvalues with increasing numbers of PCs fitted.

Table 2. Characteristics of reduced rank analyses

	Fit 3	Fit 4	Fit 5	Fit 6
No. rows^A	320,580	424,046	527,530	630,981
No. elem.s^{A,B}	7.81	10.46	13.19	15.98
"Fill-in"^{C,B}	40.98	61.01	85.21	115.60
No. op.s^D	17.42	35.14	61.47	103.77
No. par.s^E	41	46	50	53
$\log \mathscr{L}^{F,G}$	-599.6	-62.8	157.5	212.6
BIC^{H,G}	-854.6	-349.0	-153.5	-117.1

^{A} in mixed model matrix, ^{B} $\times 10^6$, ^{C} non-zero elements in Cholesky factor, ^{D} operations to factor mixed model matrix, $\times 10^9$, ^{E} parameters, ^{F} maximum log likelihood, ^{G} $+373,000$, ^{H} Bayesian Information criterion, $\times -0.5$

Figure 2 displays the first five genetic PCs. Estimates of PC1 and PC2 were essentially the same for all analyses. PC1 represented a weighted sum of all measures of 'fatness', dominated by the highly variable IMF records, and with larger contributions from records on heifers or steers. PC2
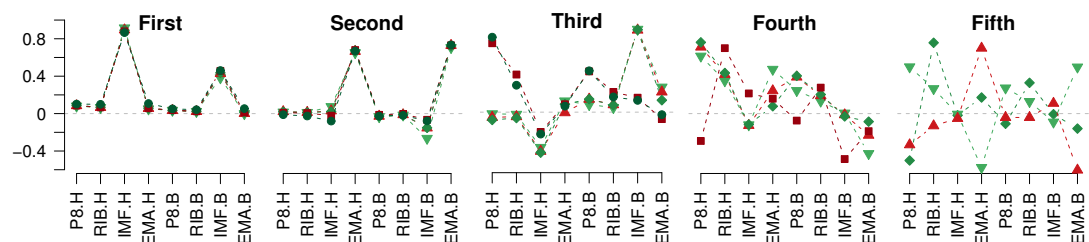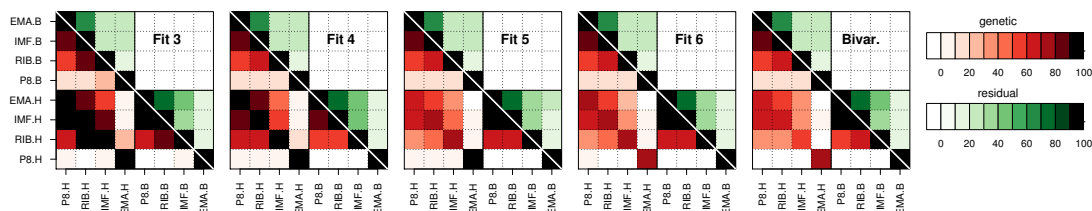

Figure 2. Estimates of genetic eigenvectors fitting 3 (●), 4 (■), 5 (♦) and 6 (▲) principal components, and bivariate analyses (▼).

58

Figure 3. Estimates of genetic (below diagonal) and residual (above diagonal) correlations ($\times 100$) for analyses fitting the first 3, 4, 5 and 6 genetic eigenvalues only, and pooled bivariate analyses.

was essentially the weighted sum EMA measurements, with a small, negative weight on IMF.B. Estimates of the third and fourth PC differed between analyses fitting only 3 or 4 PCs and the remainder, emphasizing that a least 5 PCs were required to characterise genetic covariances among the 8 traits adequately. PC3 represented the weighted difference between sexes in genetic values for the 3 'fatness' traits. PC4 and PC5 were less readily interpretable, comprising the difference between sexes for EMA, but also high weights for measurements of fat depths.

Table 3. Estimates of phenotypic variances (σ_P^2) and genetic parameters[A] from analysis fitting 6 PCs

		σ_P^2	Heifers				Bulls			
			P8	RIB	IMF	EMA	P8	RIB	IMF	EMA
H	P8	4.398	**48**	68	28	28	0	0	0	0
	RIB	2.249	86	**45**	30	26	0	0	0	0
	IMF	219.8	58	65	**36**	19	0	0	0	0
	EMA	28.53	18	20	18	**35**	0	0	0	0
B	P8	2.018	70	60	28	-1	**35**	73	41	21
	RIB	0.846	62	72	35	2	92	**28**	37	19
	IMF	120.4	40	44	72	3	66	71	**22**	16
	EMA	41.9	2	-1	-1	82	-8	-2	-1	**28**

[A] heritabilities (in bold) on, genetic correlations below, and residual correlations above diagonal; all $\times 100$

Estimates of correlations between traits are shown in Figure 3. Clearly, fitting 3 or 4 PCs only resulted in increased estimates of genetic correlations, in particular for the same trait measured on different sexes. Table 3 gives estimates of phenotypic variances and genetic parameters of the analysis fitting 6 PCs. Overall, estimates are very consistent and show good agreement with literature values. As reported previously (e.g. Meyer and Graser 1999), genetic correlations between sexes for the 'fatness' traits were only about 0.7, and records on heifers or steers were more variable and heritable than those on bulls.

CONCLUSIONS

Genetic effects for the eight traits recorded by live ultrasound scanning of beef cattle can be summarised by 5 or 6 genetic principal components. Reduced rank estimation can result in a substantial reduction in computational demands, both for genetic evaluation and variance component estimation.

REFERENCES

Henshall, J. M. and Meyer, K. (2002) *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* Comm. No. 28–12.

Kirkpatrick, M. and Meyer, K. (2004) *Genetics* **168**:2295.

Mäntysaari, E. A. (1999) In: *50th Annual Meeting*. Europ. Ass. Anim. Prod. Mimeo., 8pp.

Meyer, K. and Graser, H. U. (1999) *Proc. Ass. Advan. Anim. Breed. Genet.* **13**:385.

Meyer, K. and Kirkpatrick, M. (2005) *Genet. Select. Evol.* **37**:1.