**July 26, 2006**

Running head : PRINCIPAL COMPONENTS FOR CARCASS TRAITS

# Multivariate analyses of carcass traits for Angus cattle fitting reduced rank and factor-analytic models

Karin Meyer

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

E-mail : kmeyer@didgeridoo.une.edu.au
Phone : +61 2 6773 3331
Fax : +61 2 6773 3266

[1]AGBU is a joint venture between the NSW Department of Primary Industries and the University of New England

**Abstract**

Multivariate analyses of carcass traits for Angus cattle, consisting of 6 traits recorded on the carcass and 8 auxiliary traits measured by ultra-sound scanning of live animals, are reported. Analyses were carried out by restricted maximum likelihood, fitting a number of reduced rank and factor-analytic models for the genetic covariance matrix. Estimates of eigen-values and -vectors for different orders of fit are contrasted and implications on the estimates of genetic variances and correlations are examined.

Results indicated that at most 8 principal components were required to model the genetic covariance structure among the 14 traits. Selection index calculations suggested that the first 7 of these PCs sufficed to obtain estimates of breeding values for the carcass traits without loss in expected accuracy of evaluation. This implied that the number of effects fitted in genetic evaluation for carcass traits can be halved by estimating breeding values for the leading principal components directly.

# 1 Introduction

Characteristics of carcass quality are of considerable importance in genetic improvement programmes for beef cattle. Hence, genetic evaluation schemes, such as BREEDPLAN in Australia (Graser *et al.*, 2005), generally provide estimated breeding values (EBVs) for several carcass traits. As true carcass measurements can only be obtained at slaughter, most information contributing to genetic evaluation of stud animals is provided by corresponding traits measured on live animals via ultra-sound scanning. Generally, these traits are of little interest in their own right, and corresponding EBVs are not published.

BREEDPLAN comprises 22 and more traits in a multivariate genetic evaluation scheme. Of these, 14 represent carcass measures, but only 6 EBVs for carcass traits *per se* are reported. The other traits, measured on young, live animals, represent 3 measures of 'fatness' and a measure of lean meat yield. As heifers and steers have different patterns of growth and protein deposition than bulls, variances and heritabilities for the same measure on animals of different sex differ and within-trait, between-sex genetic correlations are less than unity (Meyer and Graser, 1999). Hence, records on different sexes (heifers and steers *versus* bulls) are treated as different traits, yielding a total of 8 auxiliary carcass traits in the analysis. Genetic correlations between several of these traits are moderately high to high, 0.7 or above, in particular for 'fatness' characteristics recorded on the same animal and between records for the same measure on different sexes.

Estimation of genetic parameters for carcass traits *per se* has been hampered by the small number of records available. Conversely, computational requirements and problems with reliable estimation of large numbers of covariance components simultaneously have severely limited multivariate analyses of all traits together. For instance, Reverter *et al.* (2000) employed 96 tri-variate, two four-variate and one six-variate analysis to estimate the complete genetic covariance matrix for the 14 carcass measures in BREEDPLAN. However, recent improvements in computing hardware available together with advances in methodology to model higher dimensional data more parsimoniously (Kirkpatrick and Meyer, 2004) and more stable algorithms for restricted maximum likelihood (REML) estimation (Meyer, 2006*a*), have made such analyses feasible.

Any set of correlated traits can be transformed into a set of new variables which are linear combinations of the original traits, are uncorrelated, and successively explain a maximum amount of variation. These new variables are generally referred to as principal components (PC), and are commonly reported in descending order of the amount of variation attributed to them. For highly correlated traits, the first few PCs then explain the bulk of variation, while the remaining PCs explain little or almost no variation. This implies that they provide virtually no information which is not already contained in the leading PCs, and that they can be ignored. This is the principle underlying the use of PCs as a dimension reduction technique. Recently, application at a genetic level has been suggested to reduce the number of EBVs to be estimated and the number of parameters to model the genetic covariance structure (Kirkpatrick and Meyer, 2004).

Earlier analyses showed that 5 to 6 PCs were required to model the genetic covariance structure among the 8 scan traits, but that the first 3 or 4 of these sufficed to summarise genetic differences between animals, explaining 97% (3 PCs) to 98.7% (4 PCs) of genetic variation (Meyer, 2005*a*). This paper presents REML estimates of genetic parameters for carcass characteristics from multivariate analyses considering all 14 traits simultaneously, fitting reduced rank and factor analytic (FA) models. In addition, the scope for dimension reduction in genetic evaluation for carcass traits of beef cattle by considering the leading principal components only is examined.

# 2 Material and methods

## 2.1 Data

Data consisted of records for carcass characteristics of Australian Angus cattle, extracted from the National Beef Recording Scheme data base in May 2005. These comprised 6 carcass traits, recorded at slaughter, and records for 4 correlated measures, recorded on live animals by ultra-sound scanning. Traits recorded on the carcass (C.) as well as on live animals were eye muscle area (EMA), intra-muscular fat content (IMF), and fat depth at the 12/13th rib (RIB) and the P8 rump site (P8). In addition, carcass weight (C.WT) and

78 percentage retail beef yield (C.RBY) were recorded at slaughter.

79 The majority of carcass measures were collected from abattoirs under a meat quality re-
80 search project by the Australian Co-operative Research Centre for Cattle and Beef Indus-
81 try conducted between 1994 and 1999; see Reverter *et al.* (2000) for details. Following
82 Reverter *et al.* (2000), records for C.EMA taken by real time ultra-sound up to 14 days
83 prior to slaughter were accepted as 'carcass' measurements. Additional records for C.WT,
84 C.RIB and C.P8 from a small number of progeny testing herds, predominantly taken from
85 2000 onwards, were available and included in the analysis. All carcass traits were recorded
86 on heifers or steers.

87 Scan traits were recorded in the field by accredited operators. As is standard practice in
88 genetic evaluation of beef cattle in Australia (Graser *et al.*, 2005), records for heifers or
89 steers (H.) and bulls (B.) were treated as separate traits. Only scan records taken between
90 300 and 700 days of age were considered. Extraction of scan records was aimed at selecting
91 records for animals which had close genetic links with animals which had carcass traits
92 recorded. Hence, only scan records in their herds of origin were considered. For these
93 herds, all contemporary groups (CG) which contained progeny of sires of animals with
94 carcass records were identified, and all records in these CG selected. There were too few
95 animals in the data which had C.RBY or C.IMF as well as H.IMF records to obtain reliable
96 estimates of the respective residual covariances. Hence H.IMF records for these animals
97 were eliminated. After further basic edits, this yielded 121 924 records on 30 427 animals.
98 Table 1 gives details for the 14 individual traits.

## 99 2.2 Estimation of variance components

100 Analyses fitted a simple animal model, including all pedigree information available for up
101 to five generations backwards. Prior to analyses, pedigrees were 'pruned', i.e. any parents
102 which did not contribute any information, because they had a single offspring only and no
103 records themselves, were treated as unknown. This was done recursively, yielding 15 501
104 parents to be included, i.e. a total of 45 928 animal genetic effects for each trait in the
105 model of analysis. Animals in the data were progeny of 1 024 sires and 12 727 dams.

Fixed effects fitted for scan traits were contemporary groups (CG), birth type (single *vs.* twin) and a dam age class (heifers *vs.* cows), the so-called 'heifer factor'. CG were defined as herd-sex-management group-date of recording subclasses, with CG subdivided further if the range of ages in a subclass exceeded 60 days. Furthermore, age at recording, nested within sex, and age of dam were fitted as a linear and quadratic covariables. This resulted in a 'double' correction for ages, which was considered necessary as neither the covariable nor the fixed effect classification alone accounted for all age differences.

Carcass traits were pre-adjusted prior to analyses, using the multiplicative adjustments of Reverter *et al.* (2000). C.WT records were adjusted to a slaughter age of 650 days, while the other carcass traits were standardised to a C.WT of 300 kg. Means ($\pm$ standard deviation) of pre-adjusted records were $335.4 \pm 52.8$ kg, $62.7 \pm 11.5$ cm$^2$, $5.374 \pm 1.81$ %, $65.5 \pm 3.1$ %, $12.28 \pm 5.04$ mm and $8.99 \pm 4.36$ mm for C.WT, C.EMA, C.IMF, C.RBY, C.P8 and C.RIB, respectively. The model for carcass traits then included CG as the only fixed effect. For data from the research project, CG were defined as herd of origin-kill regime-sex of animal subclasses, with the term 'kill regime' representing a combination of date of kill, abattoir, finishing regime and target market. For the progeny test records, CG were simply slaughter date-herd of origin-sex of animal subclasses.

Estimates of genetic and residual (co)variance matrices were obtained by REML from multivariate analyses considering all 14 traits. In addition to a 'standard' multivariate analysis, denoted as F14, which assumed the genetic covariance matrix ($\Sigma_G$) to be unstructured with 105 distinct covariance components, a number of analyses modelling the genetic dispersion structure more parsimoniously were carried out. On the one hand, these comprised reduced rank analyses fitting the first $m = 3, \ldots, 11$ genetic PCs only, denoted subsequently as F$m$, which yielded estimates of $\Sigma_G$ of rank $m$ and involved $m(29-m)/2$, i.e. from 39 (F3) to 99 (F11), parameters to model $\Sigma_G$. On the other hand, analyses that fitted a factor-analytic structure to $\Sigma_G$ considering $m = 1, \ldots, 6$ factors, denoted as F$m$+, were carried out. These yielded full rank estimates of $\Sigma_G$, and involved from 28 (F1+) to 83 parameters, consisting of 14 specific variances and $m(29-m)/2$ parameters given by the $m$ factors.

The residual covariance matrix was assumed to have full rank throughout. However, carcass traits were not measured for bulls, and heifers or steers with C.IMF or C.RBY records

did not have live scan H.IMF records in the data. Corresponding covariances were assumed to be zero. Thus there were only 63 non-zero, residual (co)variances to be estimated, i.e. $p = 91$ (F1+) to $p = 168$ (F14) parameters in total. Analyses were carried out using an 'average information' REML algorithm to fit the leading PCs only, as described by Meyer and Kirkpatrick (2005*a*), supplemented by expectation-maximisation steps. To accommodate a FA structure, a separate genetic effect, assumed to have a diagonal covariance matrix – with elements representing the 14 'specific' variances – was fitted in addition to the $m$ factors (PCs) considered, as suggested by Thompson *et al.* (2003). All calculations were carried out using our REML package WOMBAT (Meyer, 2006*b*).

Models were compared considering the REML maximum log likelihood ($\log \mathscr{L}$) and two information criteria derived from it. Akaike's information criterion ($\text{AIC}_C$), corrected for sample size, was calculated as (Burnham and Anderson, 2004)

$$AIC_C = -2\log \mathscr{L} + 2p\,(1 + (p+1)/(N-p-1))$$

where $p$ is the number of variance parameters to be estimated and $N$ denotes the total number of records in the analysis. Schwarz' Bayesian information criterion (BIC) was obtained as

$$BIC = -2\log \mathscr{L} + (N - r(\mathbf{X}))\,p$$

with $r(\mathbf{X})$ the rank of the coefficient matrix for fixed effects (Wolfinger, 1993). For our analyses, this gave a 'penalty factor' of 11.67 per parameter in the BIC.

The deviation of estimated genetic eigenvectors from an analysis fitting $m$ PCs (or factors) from those from analysis F14 was measured as the angle (in °) between corresponding vectors

$$\alpha_i = (180/\pi) \arccos\left(\mathbf{e}'_{i,m}\mathbf{e}_{i,14}/(|\mathbf{e}_{i,m}||\mathbf{e}_{i,14}|)\right)$$

where $\mathbf{e}_{i,m}$ denotes the estimate of the $i-$th eigenvector from analysis F$m$ and $|.|$ is the norm of a vector; see Kirkpatrick and Meyer (2004) for details. In addition, the similarity of estimated correlation matrices was determined as

$$\Delta r = \sum_{i=1}^{14} \sum_{j=i+1}^{14} \left(r_{ij,m} - r_{ij,14}\right)^2 / 91$$

with $r_{ij,m}$ the estimate of the correlation between traits $i$ and $j$ from an analysis fitting $m$ PCs (or factors).

## 2.3 Accuracy of genetic evaluation

The expected accuracy of genetic evaluation for carcass traits based on the first $m$ principal components only was obtained from the 'mixed' model equations (MME). These were set up for a single animal with a given amount of records or progeny information, ignoring any fixed effects, as described in the appendix (Section B.1). Assuming estimates of covariance matrices from analysis F$m$ ($m = 8, 14$), were the true values, matrices of sampling covariances among the EBVs for the $m$ principal components (and the covariances between true and estimated values) can be obtained from the inverse of the coefficient matrix in the MME, as shown, for instance, by Henderson (1975). EBVs for individual carcass traits are a linear function of the EBVs for PCs. Hence, sampling covariances for EBVs on the original scale are simple linear functions of the sampling covariances of EBVs for PCs. The expected accuracy for each trait was then simply calculated as the correlation between true and estimated breeding values.

Similarly, the MME were set up considering the first $n < m$ PCs only. This was equivalent to using an assumed genetic covariance matrix constructed from the first $n$ eigenvalues and -vectors. As this represented a scenario where the true and assumed genetic covariance matrix were different, standard formulae (Henderson, 1975) for the sampling covariances of EBVs did not apply any longer and needed to be modified, as shown in the appendix (Section B.2).

Calculations of expected accuracies were accompanied by a simple simulation study. This

involved sampling of genetic PCs and environmental effects for each trait from appropriate uni- and multivariate Normal distributions for 2000 unrelated animals, to generate records for individual traits. EBVs for PCs were obtained by setting up and solving the MME for $n = 1,\ldots,m$ PCs in turn. Transforming true and estimated breeding values for PCs to those for the individual traits, accuracies were obtained as correlations between the true and estimated values across animals. 5000 replicates were carried out, obtaining means and empirical standard deviations of accuracies across replicates.

# 3 Results

Maximum log likelihood ($\log\mathscr{L}$) values for the different analyses are listed in Table 2, together with the corresponding information criteria. For the reduced rank analyses, $\log\mathscr{L}$ increased significantly until at least 8 PCs were fitted (F8). As is often the case, this was also the best model on the basis of the $\text{AIC}_C$, though with only a small difference in $\text{AIC}_C$ values to an analysis considering 7 (F7) PCs only. In contrast, involving a more stringent penalty for the number of parameters, BIC indicated that a reduced rank model with 5 or 6 fitted best.

Similarly, $\log\mathscr{L}$ increased and $\text{AIC}_C$ decreased consistently when increasing the number of factors considered in the FA analyses. At equal number of factors, $\log\mathscr{L}$ values were substantially higher than for reduced rank analyses, in particular for low number of factors or PCs. This was due to much less partitioning of the genetic variances not accounted for by the limited number of PCs into the residual components. The extra parameters in the FA models, the specific variances, 'picked up' a large proportion of this variation. Clearly, the most parsimonious model to describe the covariance structure among the 14 traits on the basis of the BIC was a FA model with only 2 factors (F2+), involving a total of 104 parameters.

However, estimates of the total genetic variance increased and, correspondingly, estimates of the total residual variance decreased until at least 7 or 8 PCs in a reduced rank analysis, or 5 factor in a FA model were considered. This suggested that model choice on the basis of BIC might be overzealous in this case. With low numbers of records for the carcass

traits, optimal values for BIC were likely to reflect the fact that the data available only supported accurate estimation of a relatively low number of parameters. Theory indicates that the eigenvalues of the estimate of a matrix are more variable than the population values, with large values biassed upwards and small values biassed downwards while the mean is unbiassed. Our analyses constrained estimates to be larger than an operational zero. Hence, it is plausible that analyses fitting a substantial number of PCs might yield an overestimate of the sum of eigenvalues, in particular at the genetic levels. While it is not clear how large such effect may be, it is unlikely to explain a reduction from approximately 590 to 470 (F5) or 494 (F2+).

Similarity of estimated correlations with those from a standard, multivariate correlation was gaged by the average squared deviation of estimates from analysis F$m$ or F$m$+ from their counterparts from analysis F14. Results (Table 2) show substantial differences in genetic correlations for analyses fitting low numbers of PCs (or factor), with some, considerably smaller effects on the estimates of residual correlations.

## 3.1 Estimates of principal components

Figure 1 summarises estimates of eigenvalues of estimated covariances matrices. Fitting too few PCs (or factors), the leading genetic eigenvalues tended to be underestimated. Correspondingly, residual eigenvalues in the reduced rank analyses were overestimated and phenotypic values remained relatively constant, emphasizing a repartitioning of the total variation according to the number of PCs fitted. This was especially pronounced for the first eigenvalue, $PC_1$, which required at least 6 PCs in a reduced rank analysis before attaining a stable value. Genetic variance explained by higher order PCs tended to be underestimated when they were the last PC fitted (e.g. the fifth genetic eigenvalue for analysis F5), increasing to a stable value when fitting more PCs. Similar patterns have been observed in previous reduced rank analyses and simulation studies (Meyer, 2005*a*,*b*; Meyer and Kirkpatrick, 2005*b*).

For analyses fitting FA models, trends were less clear cut. With some variation partitioned into the specific variances, estimates of residual eigenvalues were less affected. Alternate genetic, and consequently phenotypic eigenvalues tended to approach a stable value from

either too high or too low values. In particular, estimates of the second and third phenotypic eigenvalues were substantially under- and overestimated when less than 5 factors were fitted.

Estimates of the corresponding eigenvectors, i.e. the weights given to the original traits when forming the PCs, are contrasted in Figure 2 for analyses F8 and F14. For the first two PCs, there was virtually no difference in estimates from the two analyses. Previous results for 8 traits (Meyer, 2005*a*) found this to hold for analyses fitting from 3 to 8 PCs, in particular for $PC_1$. This suggested that the direction of the leading PCs was estimated correctly, even if the corresponding amount of variation explained was severely underestimated. However, for our analyses at least 6 PCs in a reduced rank analysis or 5 factors in a FA model were required for this to apply and for the angles to estimates from analysis F14 to become small (see Table 2). For $PC_1$ the discrepancy was mainly due to weights for the three IMF measurements. From the third PC onwards, there were increasing differences in estimates of PCs from different analyses. Earlier simulation work demonstrated invariably large sampling variances for estimates of the higher order PCs. Fortunately, as these are associated with the smallest eigenvalues, these tend to be unimportant (Kirkpatrick and Meyer, 2004; Meyer, 2005*a*).

The first PC was dominated by C.WT, the trait with the highest genetic variance. While their was little emphasis on C.RBY and fatness characteristics measured on bulls, $PC_1$ involved positive weights for all heifer scan traits (and B.EMA) and negative weights for the corresponding traits measured on the carcass. The second PC represented, in essence, the weighted sum of all fatness measurements. Accounting for differences in genetic variation, IMF measurements were about twice as important as the fat depth records. Similarly, the main constituent as $PC_3$ was a weighted sum of all EMA measures.

## 3.2 Estimates of variances and genetic parameters

Changes in estimated genetic variances and heritabilities with increasing number of PCs (or factors) are displayed in Figure 3, together with estimates of their approximate lower bound sampling errors. Trends in estimates closely correspond to those in the estimates of genetic eigenvalues shown in Figure 1. Underestimates of eigenvalues and the total

genetic variance when fitting too few PCs in reduced rank analyses are reflected in low estimates of variances and heritabilities, which gradually increase with the number of PCs considered. We expect a reduction in sampling variances when estimating less parameters. There is some indication of approximate sampling errors to increase from analyses F8 to F14 – where estimates have essentially reached a stable value – but effects are small and not completely consistent.

The effect of the number of PCs considered on estimates of genetic correlations is illustrated in Figure 4 for the three intra-muscular fat measures. If only one PC was fitted, all correlations would be forced a have an absolute value of unity. Fitting more and more PCs attenuates the correlations until they reach stable values. Hence, a considerable proportion of correlation estimates from analysis F4 were reduced in magnitude when fitting more PCs. As emphasized by the average squared deviation in correlations given in Table 1, there was little difference in estimates betwen analyses fitting 8 or more PCs.

This is further illustrated in Figure 5, which contrasts estimates from analyses F8 and F14 for all genetic correlations and heritabilities and their approximate sampling errors. On the whole, there was close correspondence between estimates of parameters between the two analyses, with the largest difference for the correlation between C.IMF and B.IMF, which decreased from $0.59 \pm 0.11$ (F8) to $0.46 \pm 0.18$ (F14). Considering only 8 rather than all 14 PCs, estimates of sampling errors overall were reduced only by 0.01. However, as Figure 5 shows, values for the carcass traits tended to be lower for analysis F8, with a mean reduction of 0.02 and the maximum value of 0.09 for the estimate of the correlation between C.RBY and B.IMF, which was $-0.41 \pm 0.11$ and $-0.52 \pm 0.20$ for analyses F8 and F14, respectively. On the other hand, estimated sampling errors for some parameters were slightly increased for analysis F8. In part, this was explicable by an increase in magnitude. In addition, it should be borne in mind that values are estimates, based on large sample theory, and have been derived using linear approximations of non-linear functions (see Section A). Hence, some error in estimation is plausible.

All estimates from analysis F8, together with their approximate sampling errors and estimates of the phenotypic variance are summarised in Table 2. On the whole, estimates showed reasonable agreement with the results of Reverter *et al.* (2000) and previous esti-

mates for scan traits based on about twice as much data for these traits (Meyer, 2005*a*). Obviously, with small numbers of records for the carcass traits, substantial fluctuations can occur when adding additional data, such as the progeny test records for C.WT, C.P8 and C.RIB, altering the strategy to select data or changing the procedure of analysis. While some of the extreme estimates of Reverter *et al.* (2000) were moderated, others remained more or less unchanged. For instance, our estimates of genetic correlations between carcass and heifer fat depths ranged from 0.6 to 0.8 while Reverter *et al.* (2000) reported values from 0.9 to unity. As fatness traits and EMA generally show little genetic association, our estimate of $-0.24 \pm 0.14$ for the genetic correlation between C.EMA and B.IMF seems more plausible than the value of $-0.90 \pm 0.08$ obtained in the former study. Notably, however, the estimated heritability for C.RBY remained very high ($0.75 \pm 0.11$ *versus* 0.68 previously), and the estimate of a strong residual correlation with C.WT ($0.86 \pm 0.18$) was unchanged. Differences in heritability estimates for scan traits reported on different sexes were smaller than reported previously (Meyer and Graser, 1999; Meyer, 2005*a*). In particular, heritabilities for fatness traits recorded on bulls were higher, which could be due to better than average recording practices in the small number of herds selected.

## 3.3 Accuracy of genetic evaluation

Table 4 shows the expected accuracy of genetic evaluation when considering reduced numbers of principal components, for the example of a sire with 20 offspring of each sex with records for all 4 scan traits and 5 offspring with records for the 6 carcass traits. Values assuming estimates of covariances from the full rank analysis (F14) and from the reduced rank analysis with the lowest $\text{AIC}_C$ (F8) are given. Mean accuracies from the simulation study agreed closely with expected values derived from the mixed model equations, differences being at most 0.02 except for index 2 where expected values were up to 0.06 higher, and thus have been omitted. Similarly, results for 11, 12 and 13 PCs were virtually identical to those for 10 PCs, and are not shown.

Values for individual carcass traits clearly reflect the weighting these receive in successive principal components. For instance, most information on C.WT, the trait with the highest variance, is supplied by the first PC, so that the accuracy of the corresponding EBV based

only on $PC_1$ is about 90% of that achieved when considering all PCs. Conversely, EBVs for C.EMA with low weightings in the first two PCs, did not achieve a reasonable accuracy until at least three PCs were taken into account. Only for C.RIB and C.P8 did it appear advantageous to fit more than 7 PCs.

Most selection schemes consider a weighted combination of EBVs for individual traits. Hence, the effect of the number of PCs on sampling covariances as well as variances needs to be taken into account. This can be assessed by examining accuracies of indexes. Index 1 comprised C.RBY, C.IMF, C.P8 and C.EMA with relative weights of 1, 2.661, 0.060, and $-0.072$, respectively, while Index 2 weighted C.RBY C.EMA and C.P8 in a ratio of 1 to 0.354 to 0.208. These might represent indexes to select for an export market where high marbling is desirable and the domestic market (Barwick 2006; pers. comm.). Results suggest that, at the current estimates, at least 7 PCs need to be fitted so as not to compromise genetic progress.

# 4 Discussion

Results show that the genetic dispersion structure among the 14 carcass traits considered can be modelled parsimoniously by considering a subset of the genetic principal components. In contrast to other models which achieve parsimony by forcing estimates of covariance matrices to have a certain structure, no prior assumptions on the nature of covariances between variables are required. Whether a reduced rank of factor-analytic model is preferable depends on the circumstances. Both involve virtually the same calculations, relying on the identification of the leading PCs of a covariance matrix. Principal component analysis as such is merely concerned with identification of independent variables explaining the maximum amount of variation. In contrast, the underlying concept of factor analysis is to find the factors which explain the covariances between traits. This involves fitting a latent variable model with error variances equal to the specific variances. Thus, our reduced rank or 'principal component' analyses are equivalent to analyses fitting FA models where all specific variances are assumed to be zero.

Allowing for non-zero specific variances in modelling the genetic covariance matrix re-

duces the bias in estimates of the residual components when too few factors are considered. Hence, FA models tend to fit the data better than reduced rank models, in particular for small numbers of PCs, resulting in highly parsimonious models. The resulting estimates of covariance matrices are generally of full rank. Hence, FA models appear preferable to reduced rank analyses when our main objective is estimation of the covariance structure. Estimates from analyses assuming specific variances are zero have rank equal to the number of factors fitted. Here, the assumption is that all important variation is captured by the subset of factors (or PCs) considered. This results in a set of mixed model equations of size proportional to the number of PCs rather than the number of traits. As computational requirements in mixed model analyses generally increase quadratically with the number of equations, even a small reduction in the number of PCs fitted can have a dramatic impact on the efficiency of analyses; see Meyer (2005*a*) for an example.

In determining the number of factors to be fitted, there is a trade-off between bias, when omitting important PCs, and sampling variance, when fitting additional PCs which explain negligible variation. Simulations showed good agreement between the orders of fit selected on the basis of minimum BIC and the models yielding estimates of the genetic covariance with the smallest mean square errors (Meyer, 2005*a*,*b*). Corresponding estimates of genetic correlations from analysis F8 and from models with lowest BIC, a FA model with only 2 factors overall or a reduced rank model fitting 5 PCs when considering the latter models only, by and large had overlapping confidence regions. However, consistent and fairly substantial underestimates of the total genetic variation (see Table 2) and of genetic variances for individual traits (see Figure 3) were a concern. Hence, a conservative choice of 'best' model was made on the basis of $AIC_C$, which suggested that 8 PCs or 84 parameters were required to model the genetic covariance structure among the 14 traits. Selection index calculations indicated that the first 7 of these 8 PCs sufficed to account for genetic differences between animals in a genetic evaluation scheme. This implies that by adopting a parameterisation to estimate the leading genetic PC directly, the number of effects in a mixed model analysis of the 14 carcass traits could be halved.

Individual estimates of genetic parameters were generally consistent with previous and literature results. Estimated sampling errors for correlations among carcass traits and correlations between carcass and scan traits, however, were fairly substantial. Clearly,

more data for the carcass traits *per se* is required to obtain accurate estimates of their genetic relationships with the scan traits, and thus to ensure reliable estimates of breeding values based on these auxiliary traits.

# 5 Conclusions

Multivariate analyses fitting factor-analytic models for genetic covariance matrices are appealing, and can yield parsimonious models and more accurate estimates of genetic parameters than 'standard' analyses considering matrices to be unstructured. Assuming specific variances to be zero gives reduced rank estimates which can be used to estimate the leading principal only, resulting in a dimension reduction and associated decrease in computational requirements of mixed model analyses.

# Acknowledgments

This work was supported by grant BFGEN.100B of Meat and Livestock Australia Ltd (MLA).

# References


Burnham, K. P., Anderson, D. R. (2004) Multimodel inference : Understanding AIC and BIC in model selection. *Sociol. Meth. Res.* **33**:261–304. doi: 10.1177/0049124104268644.

Graser, H. U., Tier, B., Johnston, D. J., Barwick, S. A. (2005) Genetic evaluation for the beef industry in Australia. *Austr. J. Exp. Agric.* **45**. doi: 10.1071/EA05075.

Henderson, C. R. (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**:423–447.

Kirkpatrick, M., Meyer, K. (2004) Simplified analysis of complex phenotypes : Direct estimation of genetic principal components. *Genetics* **168**:2295–2306. doi: 10.1534/genetics.104.029181.

Meyer, K. (2005*a*) Genetic principal components for live ultra-sound scan traits of Angus cattle. *Anim. Sci.* **81**:337–345.

Meyer, K. (2005*b*) Sampling behaviour of reduced rank estimates of genetic covariance functions. *Proc. Ass. Advan. Anim. Breed. Genet.* **16**:286–289.

Meyer, K. (2006*a*) PX x AI : algorithmics for better convergence in restricted maximum likelihood estimation. *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod.* Communication No. 24–15.

Meyer, K. (2006*b*) "WOMBAT" – digging deep for quantitative genetic analyses using restricted maximum likelihood. *CD-ROM Eighth World Congr. Genet. Appl. Livest. Prod.* Communication No. 27–14.

Meyer, K., Graser, H. U. (1999) Estimates of parameters for scan records of Australian beef cattle treating records on males and females as different traits. *Proc. Ass. Advan. Anim. Breed. Genet.* **13**:385–388.

Meyer, K., Kirkpatrick, M. (2005*a*) Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. *Genet. Select. Evol.* **37**:1–30. doi: 10.1051/gse:2004034.

Meyer, K., Kirkpatrick, M. (2005*b*) Up hill, down dale : quantitative genetics of curvaceous traits. *Phil. Trans. R. Soc. B* **360**:1443–1455. doi: 10.1098/rstb.2005.1681.

Reverter, A., Johnston, D. J., Graser, H.-U., Wolcott, M. L., Upton, W. H. (2000) Genetic analyses of live-animal ultrasound and abattoir carcass traits in Australian Angus and Hereford cattle. *J. Anim. Sci.* **78**:1786–1795.

Thompson, R., Cullis, B. R., Smith, A. B., Gilmour, A. R. (2003) A sparse implementation of the Average Information algorithm for factor analytic and reduced rank variance models. *Austr. New Zeal. J. Stat.* **45**:445–459. doi: 10.1111/1467-842X.00297.

Wolfinger, R. D. (1993) Covariance structure selection in general mixed models. *Comm. Stat. - Simul. Comp.* **22**:1079–1106.

# A Approximation of sampling variances

At convergence, the inverse of the average information matrix gives estimates of the lower bound sampling covariances among the parameters estimated. For reduced rank estimation, these are the elements ($l_{ir}$) of the leading columns of the Cholesky factors (**L**) of the covariance matrices in the model, $\Sigma = \mathbf{LL}'$. The $ij-$th covariance component, $\sigma_{ij}$, is

$$\sigma_{ij} = \sum_{r=1}^{q(i,j)} l_{ir}\, l_{jr}$$

with $q(ij) = min(i,j,t)$ and $t$ the rank which the estimate of $\Sigma$ is set to have. The covariance between two covariances, $\sigma_{ij}$ and $\sigma_{km}$ is then

$$\mathrm{Cov}\left(\sigma_{ij},\sigma_{kl}\right) = \sum_{r=1}^{q(i,j)}\sum_{s=1}^{q(k,m)} \mathrm{Cov}\left(l_{ir}l_{j}r, l_{ks}l_{ms}\right)$$

Using a first order Taylor series expansion to approximate the product of two variables,

$$\begin{aligned}\mathrm{Cov}\left(\sigma_{ij},\sigma_{kl}\right) \approx \sum_{r=1}^{q(i,j)}\sum_{s=1}^{q(k,m)} \big[ & l_{jr}l_{ms}\,\mathrm{Cov}(l_{ir},l_{ks}) + l_{jr}l_{ks}\,\mathrm{Cov}(l_{ir},l_{ms}) \\ & + l_{ir}l_{ms}\,\mathrm{Cov}\left(l_{jr},l_{ks}\right) l_{ir}l_{ks}\,\mathrm{Cov}\left(l_{jr},l_{ms}\right)\big]\end{aligned} \tag{1}$$

Eq. 1 extends readily to two covariance components belonging to different covariance matrices, $\Sigma_1$ and $\Sigma_2$, and their respective Cholesky factors, and to the case where we parameterise to logarithmic values of the diagonal elements, $l_{ii}$. This yields an approximation to the sampling covariance matrix among all covariance components in the model of analysis, which forms the basis for calculating approximate sampling errors of genetic parameters. Again, for non-linear functions, a first order Taylor series is utilised to obtain a linear approximation whose variance is readily obtained. For a variance ratio this gives

$$\mathrm{Var}\left(\sigma_1^2/\sigma_2^2\right) \approx \left[\sigma_2^4\,\mathrm{Var}(\sigma_1^2) + \sigma_1^4\,\mathrm{Var}(\sigma_2^2) - 2\sigma_1^2\sigma_2^2\,\mathrm{Cov}(\sigma_1^2,\sigma 2^2)\right]/\sigma_2^8 \tag{2}$$

449 Similarly, for a correlation

$$\begin{aligned}\mathrm{Var}\left(\sigma_{12}/\sqrt{\sigma_1^2\sigma_2^2}\right) \approx \Big[ & 4\sigma_1^4\sigma_2^4\mathrm{Var}(\sigma_{12}) + \sigma_{12}^2\sigma_2^4\mathrm{Var}(\sigma_1^2) + \sigma_{12}^2\sigma_1^4\mathrm{Var}(\sigma_2^2) \\ & - 4\sigma_{12}\sigma_1^2\sigma_2^4\mathrm{Cov}(\sigma_{12},\sigma_1^2) - 4\sigma_{12}\sigma_1^4\sigma_2^2\mathrm{Cov}(\sigma_{12},\sigma_2^2) \\ & + 2\sigma_{12}^2\sigma_1^2\sigma_2^2\mathrm{Cov}(\sigma_1^2,\sigma_2^2)\Big] / \left(4\sigma_1^6\sigma_2^6\right) \end{aligned} \tag{3}$$

450 # B Accuracy of genetic evaluation

451 ## B.1 Mixed model equations for reduced rank covariance matrices

452 Consider the simple linear model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \tag{4}$$

453 with $\mathbf{y}$, $\mathbf{b}$, $\mathbf{u}$ and $\mathbf{e}$ denoting the vectors of observations for $k$ traits, fixed effects, random
454 effects and residuals, and $\mathbf{X}$ and $\mathbf{Z}$ the incidence matrices pertaining to $\mathbf{b}$ and $\mathbf{u}$, respec-
455 tively. Assume $\mathbf{u}$ represents a vector of genetic effects with $\mathrm{Var}(\mathbf{u}) = \mathbf{G} = \Sigma \otimes \mathbf{A}$, and $\mathbf{A}$ the
456 numerator relationship matrix, and let $\mathrm{Var}(\mathbf{e}) = \mathbf{R}$.

457 Let $\Sigma = \mathbf{E}\Lambda\mathbf{E}'$ denote the eigenvalue decomposition of the matrix of genetic covariances,
458 with $\Lambda$ the diagonal matrix of eigenvalues, $\lambda_i$, and $\mathbf{E}$ the corresponding matrix of eigenvec-
459 tors, $\mathbf{e}_i$ with $\mathbf{EE}' = \mathbf{I}$. Assume $\lambda_i$ and $\mathbf{e}_i$ are ordered in descending order of magnitude of
460 $\lambda_i$. Reparameterising (Eq. 4) to

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Z}^\star\mathbf{u}^\star + \mathbf{e} \qquad \text{with} \qquad \mathbf{Z}^\star = \mathbf{Z}(\mathbf{E} \otimes \mathbf{I}) \tag{5}$$

461 yields an equivalent model, which fits genetic values for the principal components, $\mathbf{u}^\star =$

462 $(\mathbf{E}' \otimes \mathbf{I})$, instead of the original traits. The mixed model equations for (Eq. 5) are

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}^{\star} \\ \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{Z}^{\star} + \Lambda^{-1} \otimes \mathbf{A}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}}^{\star} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^{\star\prime}\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{6}$$

463 To consider only the leading $m$ genetic principal components, replace $\mathbf{E}$ with $\mathbf{E}_m$, the $k \times m$
464 matrix comprising the first $m$ columns of $\mathbf{E}$, $\mathbf{e}_1$, ..., $\mathbf{e}_m$. This gives $\mathbf{Z}^{\star}$ with number of
465 columns proportional to $m$ rather than $k$. The number of equations in (Eq. 6) is reduced cor-
466 respondingly (replacing $\Lambda$ by its submatrix $\Lambda_m$ consisting of the first $m$ rows and columns),
467 and $\mathbf{u}^{\star}$ contains $m$ elements for each individual (Kirkpatrick and Meyer, 2004; Meyer and
468 Kirkpatrick, 2005*a*). Genetic values for the $k$ original traits can be obtained as simple
469 linear combinations of the $m$ genetic principal components,

$$\hat{\mathbf{u}} = (\mathbf{E}_m \otimes \mathbf{I})\,\hat{\mathbf{u}}^{\star} \tag{7}$$

470 Assuming that $\Sigma_m = \mathbf{E}_m \Lambda_m \mathbf{E}_m'$ is the true genetic variance matrix, i.e. that $\lambda_{m+1}, \ldots, \lambda_k$
471 are zero,

$$\mathrm{Var}\left(\hat{\mathbf{u}}^{\star}\right) = \mathrm{Cov}\left(\hat{\mathbf{u}}^{\star}, \mathbf{u}^{\star}\right) = \Lambda_m \otimes \mathbf{A} - \mathbf{C} \tag{8}$$

$$\mathrm{Var}\left(\hat{\mathbf{u}}\right) = \mathrm{Cov}\left(\hat{\mathbf{u}}, \mathbf{u}\right) = \Sigma_m \otimes \mathbf{A} - (\mathbf{E}_m \otimes \mathbf{I})\,\mathbf{C}\left(\mathbf{E}_m' \otimes \mathbf{I}\right) \tag{9}$$

472 where $\mathbf{C}$ is the part of the inverse of the coefficient matrix in (Eq. 6) pertaining to $\mathbf{u}^{\star}$.

473 Alternatively, we may have $n \leq k$ principal components with non-zero variance, but may
474 want to examine sampling (co)variances of $\hat{\mathbf{u}}$ resulting from considering the first $m < n$
475 components only. In this case, we need to distinguish between true genetic covariances,
476 $\mathbf{G} = \mathbf{E}_n \Lambda_n \mathbf{E}_n' \otimes \mathbf{A}$, and assumed values, $\tilde{\mathbf{G}} = \Sigma_m \otimes \mathbf{A}$, as outlined below (Section B.2).

## 477 B.2 True and assumed genetic covariances are different

478 In genetic evaluation via Best Linear Unbiased Prediction (BLUP), it is generally assumed
479 that the values of covariance components due to random effects and residuals are known,

480 i.e. are the population values. Under this assumption (e.g. Henderson, 1975),

$$\mathrm{Var}(\hat{\mathbf{u}}) = \mathrm{Cov}(\hat{\mathbf{u}}, \mathbf{u}) = \mathbf{G} - \mathbf{C} \qquad \text{and} \qquad \mathrm{Var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C} \tag{10}$$

481 where $\mathbf{u}$ denotes a vector of genetic values with $\mathrm{Var}(\mathbf{u}) = \mathbf{G}$, $\hat{\mathbf{u}}$ represents its best linear
482 unbiased predictor, and $\mathbf{C}$ is the part of the inverse of the coefficient matrix in the mixed
483 model equations pertaining to $\mathbf{u}$. If $\tilde{\mathbf{G}} \neq \mathbf{G}$ is used in setting up the coefficient matrix,
484 following the derivations of Henderson (1975), (co)variances in (Eq. 10) become

$$\mathrm{Var}(\hat{\mathbf{u}}) = \mathbf{G} - \mathbf{C}\tilde{\mathbf{G}}^{-1}\mathbf{G} + \left(\mathbf{I} - \mathbf{C}\tilde{\mathbf{G}}^{-1}\right)\left(\mathbf{I} - \mathbf{G}\tilde{\mathbf{G}}^{-1}\right)\mathbf{C} \tag{11}$$

$$\mathrm{Cov}(\hat{\mathbf{u}}, \mathbf{u}) = \mathbf{G} - \mathbf{C}\tilde{\mathbf{G}}^{-1}\mathbf{G} \tag{12}$$

$$\mathrm{Var}(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{C} + \mathbf{C}\left(\mathbf{I} - \tilde{\mathbf{G}}^{-1}\mathbf{G}\right)\tilde{\mathbf{G}}^{-1}\mathbf{C} \tag{13}$$

Table 1: Characteristics of the data for traits measured on the carcass (C.WT : weight, C.EMA : eye muscle area, C.P8 : rump fat, C.RIB : rib fat, C.IMF : intra-muscular fat, and C.RBY : retail beef yield), and measured on live heifers or steers (H.EMA : eye muscle area, H.P8 rump fat, H.RIB rib fat, and H.IMF intra-muscular fat) and live bulls (B.EMA : eye muscle area, B.P8 : rump fat, B.RIB : rib fat, and B.IMF : intra-muscular fat).

| Trait | Number of records | Mean | Standard deviation | Minimum | Maximum | Mean age (days) | Number of CG[a] |
|---|---|---|---|---|---|---|---|
| C.WT (kg) | 3 780 | 348.9 | 82.8 | 157 | 518 | 696.9 | 305 |
| C.RBY (%) | 883 | 67.0 | 3.7 | 54 | 76 | – | 145 |
| C.EMA ($cm^2$) | 1 847 | 63.4 | 10.3 | 34 | 110 | – | 232 |
| C.P8 (mm) | 3 385 | 15.34 | 8.57 | 1 | 36 | – | 291 |
| C.RIB (mm) | 2 640 | 9.77 | 4.94 | 1 | 29 | – | 273 |
| C.IMF (%×10) | 1 490 | 47.8 | 20.0 | 12 | 127 | – | 234 |
| H.EMA ($cm^2$) | 18 170 | 59.1 | 9.1 | 27 | 97 | 508.2 | 478 |
| H.P8 (mm) | 18 362 | 6.34 | 3.15 | 1 | 25 | 507.5 | 481 |
| H.RIB (mm) | 18 278 | 4.88 | 2.36 | 1 | 20 | 507.1 | 482 |
| H.IMF (%×10) | 14 276 | 45.2 | 20.3 | 2 | 86 | 519.4 | 333 |
| B.EMA ($cm^2$) | 10 409 | 73.6 | 11.9 | 30 | 115 | 468.6 | 399 |
| B.P8 (mm) | 10 313 | 3.79 | 1.81 | 1 | 21 | 469.0 | 395 |
| B.RIB (mm) | 10 405 | 3.06 | 1.44 | 1 | 19 | 468.7 | 399 |
| B.IMF (%×10) | 7 686 | 25.3 | 16.2 | 1 | 82 | 474.0 | 284 |

[a] contemporary groups

Table 2: Number of parameters ($p$) for different analyses (F$n$ : analysis fitting the leading $n$ principal components, F$n$+ : analysis fitting a factor-analytic model with $n$ factors), together with the maximum log likelihood ($\log \mathscr{L}$) values and Akaike ($\text{AIC}_C$) and Bayesian (BIC) information criteria (all scaled as deviation from the respective 'best' values), estimates of the total variation ($\sum_i \lambda_i$) and measures of discrepancy to estimates from analysis F14 ($\sqrt{\Delta r}$ : square root of the average squared deviation of correlations, $\alpha_i$ : angle between estimates of the $i-$th eigenvectors).

| | $p$ | $\log \mathscr{L}$ | -½$\text{AIC}_C$ | -½BIC | genetic | | | | | residual | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $\sum_i \lambda_i$ | $\sqrt{\Delta r}$ | $\alpha_1$ | $\alpha_2$ | $\alpha_3$ | $\sum_i \lambda_i$ | $\sqrt{\Delta r}$ |
| F3 | 102 | -407.3 | -357.3 | -244.7 | 386.6 | 0.273 | 37.3 | 37.0 | 61.2 | 953.3 | 0.044 |
| F4 | 113 | -152.4 | -113.5 | -54.0 | 402.8 | 0.227 | 39.7 | 40.1 | 47.8 | 937.5 | 0.043 |
| F5 | 123 | -82.2 | -53.2 | -42.1 | 471.3 | 0.190 | 29.0 | 28.9 | 70.4 | 874.3 | 0.054 |
| F6 | 132 | -30.2 | -10.3 | -42.7 | 566.6 | 0.093 | 4.0 | 3.8 | 23.1 | 792.5 | 0.045 |
| F7 | 140 | -14.0 | -2.0 | -73.1 | 588.2 | 0.068 | 1.7 | 2.0 | 20.6 | 772.9 | 0.020 |
| F8 | 147 | -4.9 | 0 | -104.9 | 593.1 | 0.035 | 0.8 | 1.5 | 23.6 | 768.2 | 0.009 |
| F9 | 153 | -0.6 | -1.7 | -135.6 | 594.8 | 0.010 | 0.8 | 0.8 | 3.2 | 765.7 | 0.004 |
| F10 | 158 | -0.1 | -6.2 | -164.3 | 592.9 | 0.005 | 0.2 | 0.2 | 1.1 | 767.0 | 0.001 |
| F11 | 162 | 0.0 | -10.1 | -187.5 | 592.6 | 0.001 | 0.1 | 0.1 | 0.7 | 767.2 | 0.000 |
| F14 | 168 | -0.0 | -16.2 | -222.5 | 590.6 | 0 | 0 | 0 | 0 | 768.7 | 0 |
| F1+ | 91 | -276.7 | -215.7 | -49.9 | 440.7 | 0.232 | 89.3 | 0.0 | 0.0 | 886.2 | 0.054 |
| F2+ | 104 | -150.9 | -103.0 | 0 | 494.1 | 0.183 | 18.4 | 27.5 | 0.0 | 843.6 | 0.043 |
| F3+ | 116 | -98.6 | -62.6 | -17.6 | 538.0 | 0.134 | 26.6 | 29.7 | 35.7 | 807.0 | 0.038 |
| F4+ | 127 | -53.8 | -28.8 | -37.1 | 546.0 | 0.117 | 30.9 | 31.9 | 22.1 | 805.2 | 0.046 |
| F5+ | 137 | -24.0 | -9.1 | -65.6 | 589.1 | 0.085 | 7.1 | 6.5 | 10.8 | 770.5 | 0.027 |
| F6+ | 146 | -10.6 | -4.7 | -104.7 | 597.0 | 0.046 | 1.2 | 1.7 | 12.1 | 764.6 | 0.022 |

Table 3: Estimates of genetic parameters (×100; heritabilities on diagonal (in bold), genetic correlations below diagonal, and residual correlations above diagonal), together with their approximate lower bound sampling errors (×100), and estimates of phenotypic variances ($\sigma^2_P$), from reduced rank analysis fitting 8 principal components (see Table 1 for abbreviations of trait names).

| | Carcass | | | | | | Heifers/steers | | | | Bulls | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C.WT | C.RBY | C.EMA | C.P8 | C.RIB | C.IMF | H.EMA | H.P8 | H.RIB | H.IMF | B.EMA | B.P8 | B.RIB | B.IMF |
| C.WT | **51**±6 | 86±18 | -8±5 | -19±6 | -23±5 | -22±9 | 28±5 | 5±5 | 10±5 | -5±6 | – | – | – | – |
| C.RBY | 10±13 | **75**±11 | -21±11 | -32±11 | -14±9 | -33±15 | 39±14 | -7±12 | 3±12 | – | – | – | – | – |
| C.EMA | -46±9 | 23±16 | **22**±4 | 22±4 | 23±4 | 15±6 | 52±3 | 21±4 | 20±4 | 16±6 | – | – | – | – |
| C.P8 | -18±9 | -52±10 | -3±13 | **38**±5 | 36±3 | 16±7 | 9±4 | 30±4 | 22±4 | 20±5 | – | – | – | – |
| C.RIB | -18±11 | -82±8 | -21±14 | 83±7 | **26**±4 | 18±6 | 2±4 | 16±4 | 23±4 | 8±5 | – | – | – | – |
| C.IMF | -30±8 | -43±10 | -21±12 | 26±8 | 31±10 | **58**±5 | -7±6 | 4±7 | 15±6 | – | – | – | – | – |
| H.EMA | 51±7 | 1±11 | 47±9 | -4±8 | -11±10 | -36±6 | **31**±2 | 30±2 | 29±2 | 20±2 | – | – | – | – |
| H.P8 | 17±8 | -53±11 | -18±10 | 77±7 | 73±9 | 28±6 | 19±5 | **41**±2 | 71±1 | 35±2 | – | – | – | – |
| H.RIB | 19±8 | -56±12 | -28±11 | 62±8 | 78±9 | 22±6 | 18±6 | 87±5 | **36**±3 | 38±1 | – | – | – | – |
| H.IMF | 33±8 | -42±10 | -32±11 | 25±8 | 32±10 | 69±4 | 21±5 | 58±4 | 62±5 | **31**±2 | – | – | – | – |
| B.EMA | 43±10 | 41±11 | 56±11 | -23±9 | -36±11 | -44±7 | 87±5 | 0±6 | -4±7 | 1±7 | **26**±3 | 25±2 | 25±2 | 21±2 |
| B.P8 | -2±12 | -62±12 | -19±14 | 63±10 | 81±12 | 34±9 | -4±8 | 70±6 | 64±9 | 32±7 | -8±8 | **41**±4 | 69±1 | 46±2 |
| B.RIB | -9±12 | -53±12 | -12±14 | 62±9 | 82±10 | 25±9 | -4±8 | 55±8 | 68±7 | 28±8 | -6±8 | 90±5 | **37**±4 | 42±2 |
| B.IMF | 17±11 | -41±12 | -24±14 | 41±9 | 51±11 | 59±7 | 5±7 | 40±6 | 46±7 | 65±6 | 4±8 | 70±5 | 75±5 | **24**±3 |
| $\sigma^2_P$ | 620.1 | 4.584 | 33.03 | 11.36 | 7.539 | 224.0 | 31.46 | 4.320 | 2.216 | 231.0 | 42.30 | 1.812 | 0.857 | 147.1 |

Table 4: Accuracy of genetic evaluation (E : expected value, *se* : empirical standard deviation from simulation) for carcass traits (see Table 1 for abbreviations) and two selection indexes( IND1 and IND2) for a sire with 20 male and female progeny with records for the 4 live scan traits each and 5 progeny with records for the 6 carcass traits, considering increasing numbers of genetic principal components and assuming estimates from analysis F14 or F8 are the population values.

| No. PCs | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Full rank analysis (F14) | | | | | | | | | | | | |
| C.WT | E | 66.76 | 67.09 | 68.10 | 72.82 | 72.82 | 72.86 | 73.25 | 73.27 | 73.34 | 73.35 | 73.35 |
| | *se* | 1.23 | 1.22 | 1.20 | 1.05 | 1.05 | 1.05 | 1.04 | 1.04 | 1.04 | 1.04 | 1.04 |
| C.RBY | E | 5.01 | 64.04 | 70.34 | 70.75 | 73.03 | 72.40 | 81.66 | 82.00 | 82.46 | 82.53 | 82.54 |
| | *se* | 2.22 | 1.31 | 1.12 | 1.11 | 1.03 | 1.05 | 0.74 | 0.73 | 0.72 | 0.71 | 0.71 |
| C.EMA | E | 10.94 | 25.91 | 45.01 | 73.50 | 73.44 | 74.50 | 74.70 | 74.76 | 74.79 | 74.80 | 74.82 |
| | *se* | 2.18 | 2.10 | 1.77 | 1.03 | 1.03 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| C.IMF | E | 27.48 | 71.32 | 73.65 | 79.21 | 79.31 | 79.58 | 79.68 | 79.68 | 80.38 | 80.40 | 80.40 |
| | *se* | 2.09 | 1.09 | 1.02 | 0.85 | 0.85 | 0.84 | 0.83 | 0.83 | 0.81 | 0.81 | 0.81 |
| C.P8 | E | 14.88 | 45.82 | 61.06 | 62.12 | 65.38 | 73.62 | 74.52 | 77.78 | 77.93 | 78.06 | 78.06 |
| | *se* | 2.20 | 1.77 | 1.41 | 1.39 | 1.28 | 1.02 | 0.99 | 0.88 | 0.88 | 0.87 | 0.87 |
| C.RIB | E | 18.19 | 58.03 | 71.98 | 77.26 | 79.84 | 80.36 | 81.61 | 81.81 | 84.41 | 84.73 | 84.73 |
| | *se* | 2.14 | 1.49 | 1.09 | 0.91 | 0.81 | 0.80 | 0.75 | 0.75 | 0.65 | 0.64 | 0.64 |
| IND1 | E | 24.63 | 26.32 | 47.04 | 62.75 | 67.66 | 68.19 | 72.92 | 73.12 | 75.30 | 75.30 | 75.30 |
| | *se* | 2.10 | 2.05 | 1.74 | 1.36 | 1.21 | 1.19 | 1.04 | 1.04 | 0.96 | 0.96 | 0.96 |
| IND2 | E | 5.01 | 64.04 | 70.34 | 70.75 | 73.03 | 72.40 | 81.66 | 82.00 | 82.46 | 82.53 | 82.54 |
| | *se* | 2.22 | 1.31 | 1.12 | 1.11 | 1.03 | 1.05 | 0.74 | 0.73 | 0.72 | 0.71 | 0.71 |
| Reduced rank analysis (F8) | | | | | | | | | | | | |
| C.WT | E | 67.63 | 68.27 | 71.08 | 73.60 | 73.61 | 73.55 | 74.14 | 74.16 | | | |
| | *se* | 1.22 | 1.20 | 1.12 | 1.03 | 1.03 | 1.04 | 1.02 | 1.02 | | | |
| C.RBY | E | 6.49 | 57.09 | 59.97 | 62.34 | 70.23 | 71.32 | 82.21 | 82.49 | | | |
| | *se* | 2.24 | 1.51 | 1.43 | 1.37 | 1.12 | 1.09 | 0.72 | 0.71 | | | |
| C.EMA | E | 11.69 | 22.62 | 59.73 | 73.06 | 73.59 | 74.33 | 74.49 | 74.58 | | | |
| | *se* | 2.23 | 2.13 | 1.46 | 1.05 | 1.03 | 1.01 | 1.00 | 1.00 | | | |
| C.IMF | E | 27.82 | 76.63 | 76.07 | 82.51 | 84.06 | 84.22 | 84.29 | 84.29 | | | |
| | *se* | 2.06 | 0.93 | 0.95 | 0.71 | 0.66 | 0.65 | 0.65 | 0.65 | | | |
| C.P8 | E | 16.06 | 45.74 | 62.33 | 67.27 | 71.51 | 73.33 | 76.57 | 80.28 | | | |
| | *se* | 2.15 | 1.77 | 1.36 | 1.23 | 1.10 | 1.03 | 0.92 | 0.79 | | | |
| C.RIB | E | 19.31 | 58.38 | 71.69 | 82.68 | 87.78 | 87.67 | 88.26 | 88.67 | | | |
| | *se* | 2.14 | 1.48 | 1.09 | 0.70 | 0.51 | 0.51 | 0.49 | 0.47 | | | |
| IND1 | E | 24.26 | 38.72 | 39.31 | 62.47 | 76.04 | 75.65 | 80.93 | 81.13 | | | |
| | *se* | 2.11 | 1.92 | 1.91 | 1.36 | 0.95 | 0.96 | 0.78 | 0.77 | | | |
| IND2 | E | 6.49 | 57.09 | 59.97 | 62.34 | 70.23 | 71.32 | 82.21 | 82.49 | | | |
| | *se* | 2.26 | 1.53 | 1.46 | 1.40 | 1.15 | 1.09 | 0.72 | 0.72 | | | |

Figure 1: Estimates of the first 6 eigenvalues of genetic (•), residual (▲) and phenotypic (⧫) covariance matrices from a full rank, multivariate analyses (F14), reduced rank analyses fitting increasing numbers of genetic principal components (F3 to F11) and analyses fitting a factor-analytic structure for the genetic covariance matrix with increasing numbers (F1+ to F6+) of factors (Horizontal lines indicate values for analysis F14).

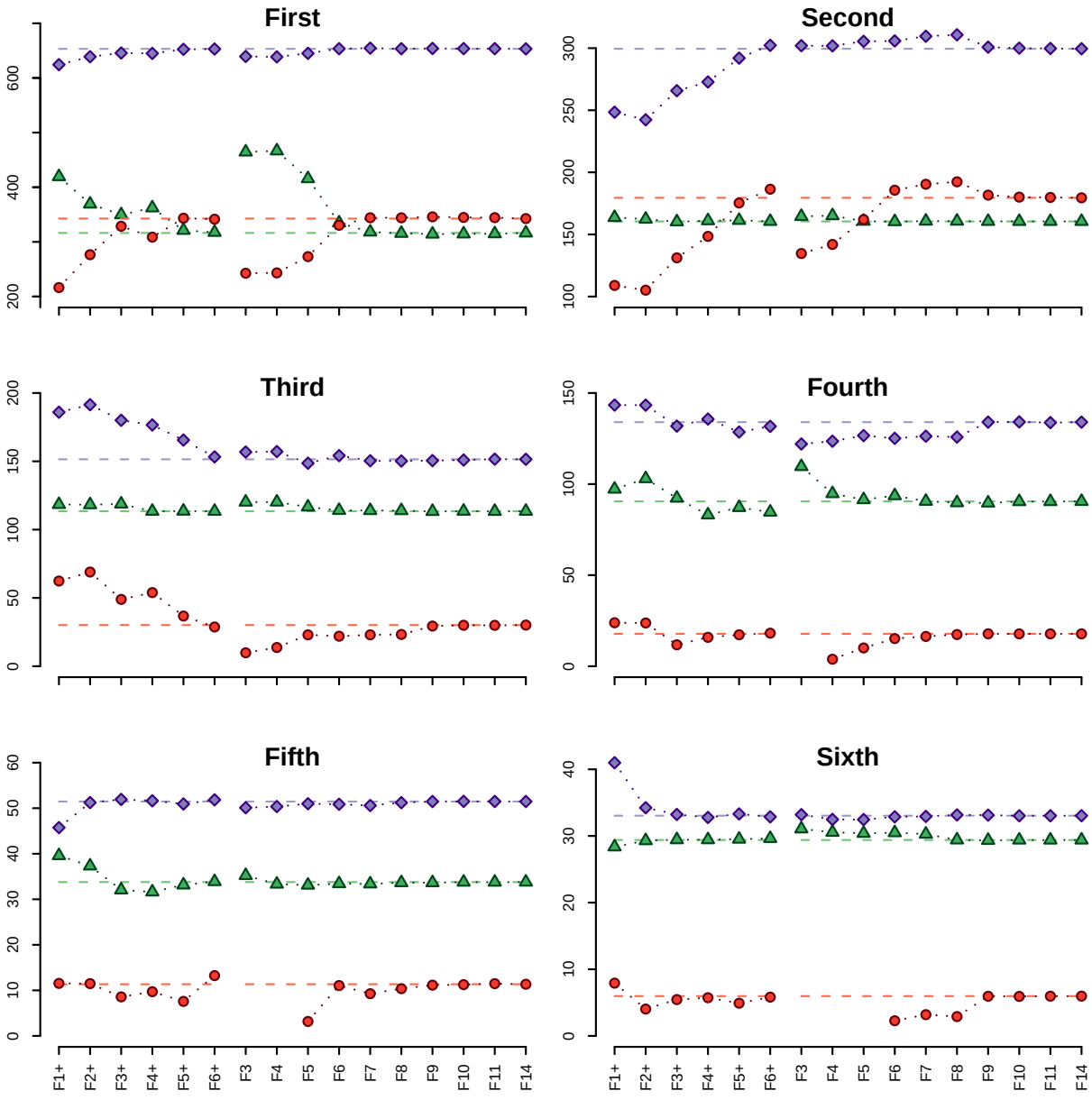

Figure 2: Estimates of weights for individual traits (see Table 1 for abbreviations) in the first 6 genetic principal components from a reduced rank analysis fitting 8 principal components (• : as estimated, and ∘ : divided by estimated genetic standard deviation and scaled by 6 for first and second and scaled by 3 otherwise), and a full rank analysis (×).

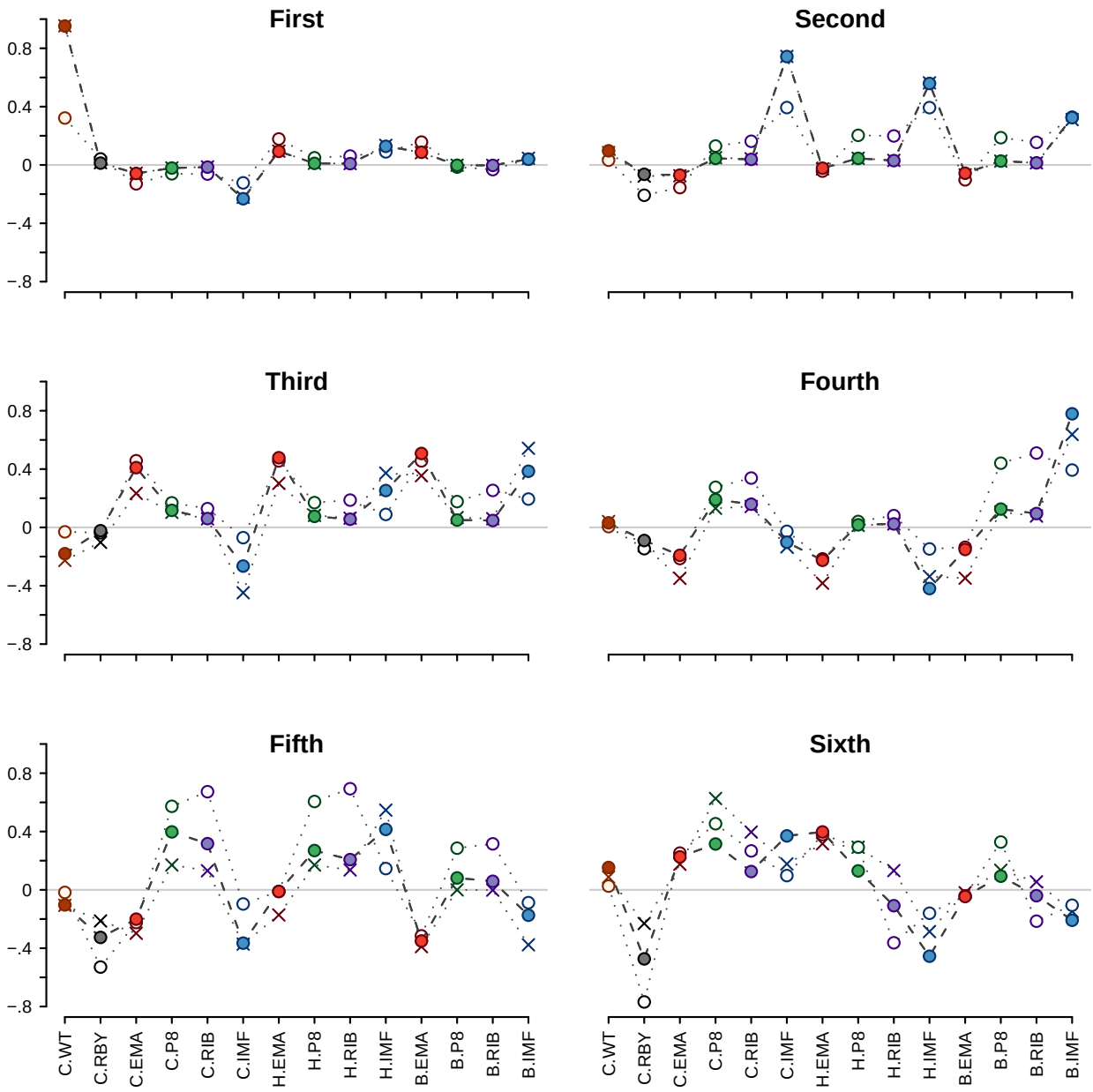

Figure 3: Estimates of genetic variances (•) and heritabilities (⧫) for selected traits (see Table 1 for definition of abbreviations), together with their approximate lower bound sampling errors, from reduced rank analyses fitting increasing numbers of principal components (F3 to F11), a standard multivariate analysis (F14) and analyses fitting a factor-analytic model (F1+ to F6+).

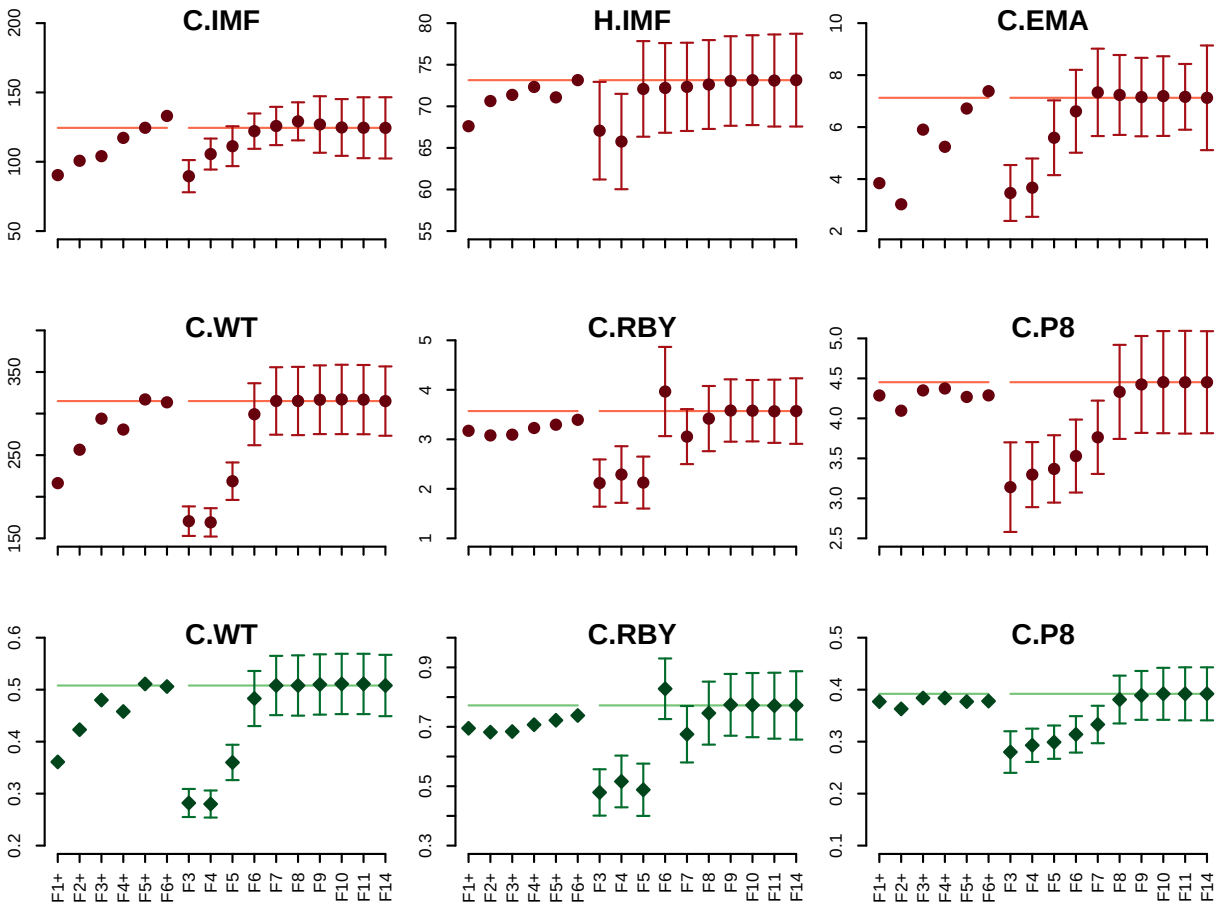

Figure 4: Estimates of genetic correlations for intra-muscular fat content measured on the carcass (left), on live heifers and steers (middle) or bulls (right) with all other traits (see Table 1 for abbreviations), from analyses fitting 4 (▼), 6 (•), 8 (▲), and all 14 (♦) genetic principal components.

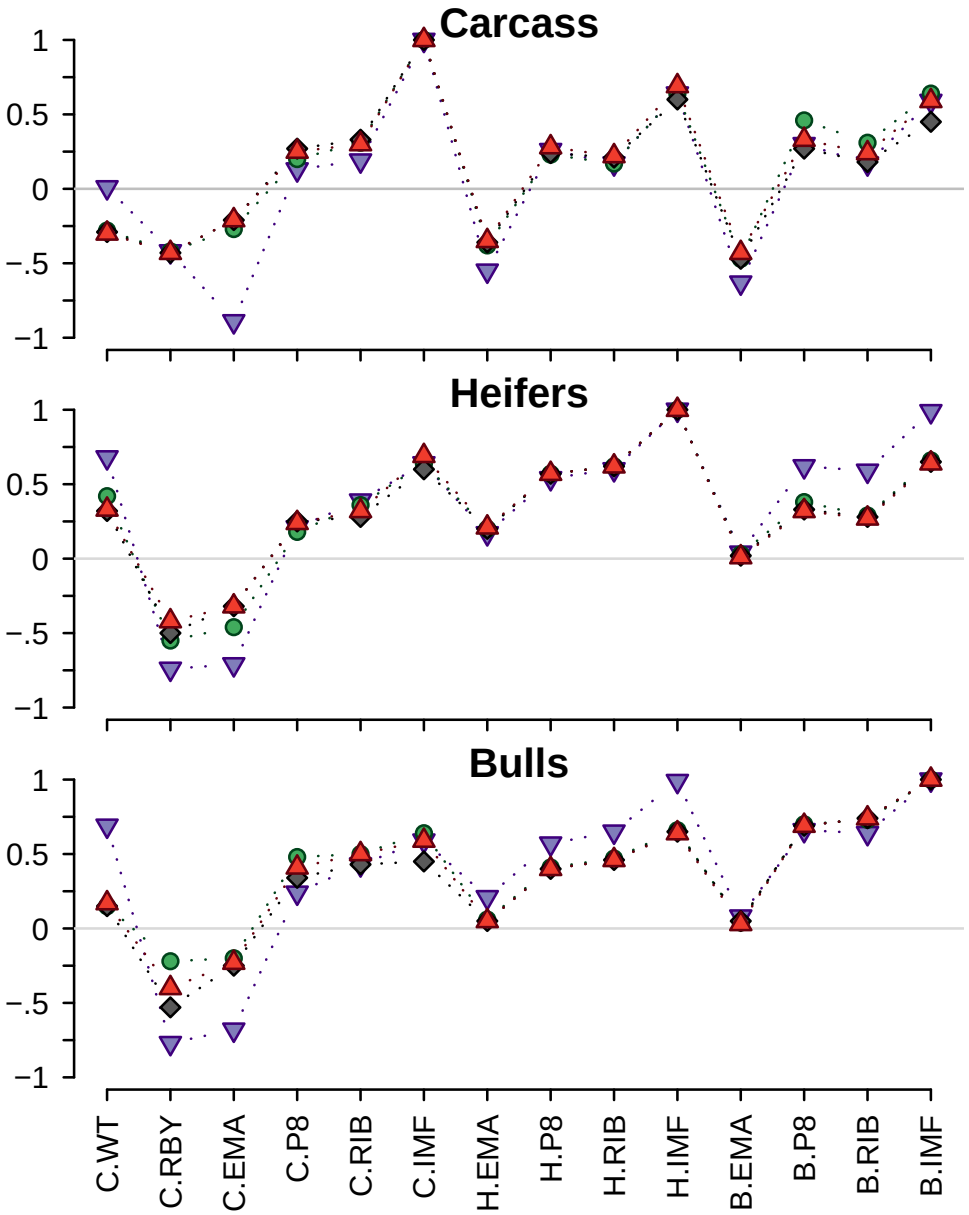

Figure 5: Estimates of genetic parameter (•) and their approximate lower bound sampling errors (⧫) from a full rank analysis (F14) and a reduced rank analysis fitting 8 principal components (F8); closed symbols pertain to parameters among the live scan traits only, open symbols pertain to estimates involving carcass traits.