

MULTIVARIATE ESTIMATION OF GENETIC PARAMETERS – QUO VADIS?

Karin Meyer

Animal Genetics and Breeding Unit*, University of New England, Armidale, NSW 2351

SUMMARY

Problems inherent in multivariate, genetic analyses to estimate covariance components are discussed. New developments in methodology with the scope to yield ‘better’ estimates are described, and their application is demonstrated for an analysis of carcass traits of beef cattle.

INTRODUCTION

Estimation of genetic parameters is one of the basic tasks in quantitative genetics. As recording schemes become more sophisticated and breeding objectives more detailed, the number of traits of interest is increasing continually. This necessitates multivariate analyses considering more than just a few traits simultaneously. Fortunately, we are at a stage where advances in modelling, computational algorithms and the corresponding software for estimation, paired with modern day computer hardware are bringing large-scale analyses comprising numerous traits and records on tens of thousands of animals within the realms of reality. For example, Tyrisevä *et al.* (2011) recently presented a 25-trait analysis involving more than 100 000 sires.

However, comparatively little attention has been paid to the problems of sampling variation inherent in multivariate analyses comprising multiple traits. It is well known that the eigenvalues of estimated covariance matrices are systematically over-dispersed (Lawley 1956) and that a large proportion of the sampling variances of genetic parameter estimates can be attributed to this excess variation. Moreover, the effects of this phenomenon increase dramatically with the number of traits. Hence, even multi-dimensional analyses based on relatively large data sets are likely to yield imprecise estimates. At the other end of the spectrum, we have numerous scenarios where the numbers of records are invariably limited. This includes records for new traits of interest or traits which are difficult or expensive to measure but which may have substantial impact on selection decisions in livestock improvement programmes. Typical examples are carcass characteristics of beef cattle. Similarly, evolutionary biologist concerned with quantitative genetics of natural populations are usually restricted to small samples.

Hence, any avenue to ‘improve’ estimates, i.e. to obtain estimates which are on average closer to the population values, should be carefully considered. On the one hand, we have accumulated a substantial body of knowledge about genetic parameters for various traits. However, typically this is completely ignored. While the Bayesian paradigm directly provides the means to incorporate such prior information, analyses concerned with the estimation of covariance components more often than not assume flat or uninformative priors (Thompson *et al.* 2005). On the other hand, statistical techniques are available – often referred to as regularization methods – which substantially reduce sampling variance, albeit at the expense of introducing some bias, and thus yield ‘better’ estimates. Interest in regularized estimation for multivariate analyses dates back to the Seventies and earlier, stimulated in particular by the work of Stein (e.g. James and Stein 1961; Stein 1975). Recently, there has been a resurgence in attention with applications for estimation in very high-dimensional settings, in particular for genomic data (e.g. Warton 2008; Yap *et al.* 2009; Witten and Tibshirani 2009).

This paper reviews the principles involved and examines the scope for adapting such techniques to estimation of genetic parameters for continuous traits in a mixed model framework. A penalized maximum likelihood scheme and suitable penalties are presented together with an application.

*AGBU is a joint venture of NSW Department of Industry and Investment and the University of New England

IMPROVED ESTIMATION

The quality of a statistical estimator is generally quantified by some measure of the difference between the estimator and the true value, or *loss*. A commonly used quantity is the mean square error. This is a quadratic loss, comprised of the sampling variance and the square of the bias in the estimator. We talk about improving an estimator when we are able to modify it in some way so that, on average, it is closer to the true value, i.e. has reduced loss. Usually this involves a trade-off between a reduction in sampling variance and additional bias.

For covariance matrices, commonly employed measures of divergence are the entropy (L_1) and quadratic (L_2) loss (James and Stein 1961):

$$L_1(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma}) - \log|\Sigma^{-1} \hat{\Sigma}| - q \quad \text{and} \quad L_2(\Sigma, \hat{\Sigma}) = \text{tr}(\Sigma^{-1} \hat{\Sigma} - \mathbf{I})^2 \quad (1)$$

where Σ and $\hat{\Sigma}$ denote a covariance matrix of size $q \times q$ and its estimator, respectively, and q represents the number of traits.

A reduction in loss can often be achieved by regularizing estimators. In broad terms, *regularization* describes a scenario where estimation for ill-posed or overparameterized problems is improved through use of some form of additional information. Often the latter is composed of a penalty for a deviation from a desired outcome. For example, in fitting smoothing splines a ‘roughness penalty’ is employed to place preference on simple, smooth functions (Green 1998). Well known forms of regularization are ridge regression (Hoerl and Kennard 1970) and the LASSO (Tibshirani 1996).

PENALIZED MAXIMUM LIKELIHOOD

Let $\log \mathcal{L}(\theta)$ denote the standard likelihood pertaining to a given model and vector of parameters θ in a maximum (ML) or restricted maximum likelihood (REML) framework of estimation. Modified estimates can be obtained by maximizing the *penalized* likelihood

$$\log \mathcal{L}_P(\theta) = \log \mathcal{L}(\theta) - \frac{1}{2} \psi \mathcal{P}(\theta) \quad (2)$$

where the penalty $\mathcal{P}(\theta)$ is a selected function of the parameters – aimed at reducing loss in their estimates – and ψ is a tuning factor which specifies the relative emphasis to be given to the penalty compared to the usual, unpenalized estimator. Penalizing the likelihood provides a direct link to Bayesian estimation: For a given prior distribution of the parameters, a corresponding penalty can be obtained as minus the logarithmic value of the density of the prior.

Penalizing eigenvalues. Recognition of the systematic upwards bias in the largest and downwards bias in the smallest eigenvalues of estimated covariance matrices early on has led to the development of various improved estimators which modify the eigenvalues in some fashion whilst retaining the corresponding eigenvectors. As the mean eigenvalue is expected to be unbiased, a specific proposal has been to regress all eigenvalues towards their mean in order to reduce their excessive spread. This is equivalent to assuming eigenvalues have a prior that is a Normal distribution. It yields an estimator that is a weighted combination of the sample covariance matrix and a multiple of the identity matrix.

Considering a one-way analysis of variance to estimate the genetic covariance matrix, Σ_G , Hayes and Hill (1981) proposed to apply the same type of shrinkage to the canonical eigenvalues, λ_i , i.e. the eigenvalues of $\Sigma_P^{-1} \Sigma_G$, with Σ_P denoting the phenotypic covariance matrix. The resulting estimate of Σ_G is a weighted combination of the standard estimate $\hat{\Sigma}_G$ and $\bar{\lambda} \hat{\Sigma}_P$, with $\bar{\lambda}$ the mean of the λ_i . The authors thus described their method as ‘bending’ Σ_G towards Σ_P , and argued that this was advantageous as Σ_P typically is estimated much more accurately than Σ_G . Hayes and Hill (1981) presented a simulation study demonstrating that this procedure could substantially increase the achieved response to selection based on an index derived using the modified estimates. This implies that ‘bending’ resulted in estimates closer to the population values than unmodified estimates.

Recently, Meyer and Kirkpatrick (2010) demonstrated that the equivalent to bending in a (RE)ML framework could be obtained by placing a penalty proportional to the variance among the estimated canonical eigenvalues on the likelihood:

$$\mathcal{P}_\lambda(\boldsymbol{\theta}) \propto \text{tr}(\boldsymbol{\Lambda} - \bar{\lambda}\mathbf{I})^2 \quad \text{with} \quad \bar{\lambda} = \text{tr}(\boldsymbol{\Lambda})/q \quad (3)$$

for $\boldsymbol{\Lambda} = \text{Diag}\{\hat{\lambda}_i\}$. They showed by simulation that this yielded a substantial reduction in loss for animal model analyses, not only for data with a paternal half-sib family structure but also for data with many different types of covariances between animals. An alternative form, $\mathcal{P}_\lambda^f(\boldsymbol{\theta})$, is obtained by penalizing the eigenvalues on the logarithmic scale, i.e. defining $\boldsymbol{\Lambda} = \text{Diag}\{\log(\hat{\lambda}_i)\}$. A disadvantage of $\mathcal{P}_\lambda(\boldsymbol{\theta})$ or $\mathcal{P}_\lambda^f(\boldsymbol{\theta})$ is that it is not readily extended to models with more than two random effects. The canonical decomposition gives $\boldsymbol{\Sigma}_G = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'$ and the residual covariance matrix, $\boldsymbol{\Sigma}_E = \mathbf{T}(\mathbf{I} - \boldsymbol{\Lambda})\mathbf{T}'$, with \mathbf{I} an identity matrix and \mathbf{T} the matrix of eigenvectors of $\boldsymbol{\Sigma}_P^{-1}\boldsymbol{\Sigma}_G$ scaled by a matrix square root of $\boldsymbol{\Sigma}_P$. Hence, $\mathcal{P}_\lambda(\boldsymbol{\theta})$ can be thought of as penalizing both $\boldsymbol{\Sigma}_G$ and $\boldsymbol{\Sigma}_E$.

Penalty $\mathcal{P}_\lambda(\boldsymbol{\theta})$ is based on the assumption that all λ_i are sampled from a distribution with common mean $\bar{\lambda}$. Hence, using $\mathcal{P}_\lambda(\boldsymbol{\theta})$ has been found to result in over-shrinkage when the corresponding population values were spread far apart, even when applying $\mathcal{P}_\lambda(\boldsymbol{\theta})$ to $\log(\lambda_i)$ rather than λ_i (Meyer and Kirkpatrick 2010). An alternative is to assume that the true λ_i are evenly distributed. As $\lambda_i \in [0, 1]$, a suitable distribution might be that of the order statistics on the unit interval. These have a Beta distribution. Treating the λ_i as independent gives a penalty

$$\mathcal{P}_\beta(\boldsymbol{\theta}) \propto \sum_{i=1}^q (i-1) \log(\lambda_i) + (q-i) \log(1-\lambda_i) \quad (4)$$

Arguing that unpenalized estimates of the extreme eigenvalues $\hat{\lambda}_q^0$ and $\hat{\lambda}_1^0$ are overdispersed, i.e. that the true values lie in the interval $[\hat{\lambda}_q^0, \hat{\lambda}_1^0]$, we may wish to apply $\mathcal{P}_\beta(\boldsymbol{\theta})$ after scaling to $(\lambda_i - \hat{\lambda}_q^0)/(\hat{\lambda}_1^0 - \hat{\lambda}_q^0)$.

Penalties on matrix divergence. A standard assumption in Bayesian estimation of a covariance matrix is that of an Inverse Wishart prior distribution, $p(\boldsymbol{\Sigma}|\boldsymbol{\Omega}, \nu) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(\nu+q+1)} \exp[-\frac{1}{2} \text{tr}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega})]$ (e.g Sorensen and Gianola 2002), with scale parameter $\boldsymbol{\Omega}$ and degree of belief ν . Omitting terms not depending on $\boldsymbol{\Sigma}$ or $\boldsymbol{\Omega}$ and taking logarithms gives $(\nu + q + 1) \log |\boldsymbol{\Sigma}| + \nu \text{tr}(\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Omega})$.

To ‘borrow strength’ from the phenotypic covariance matrix as above, a penalty which regularizes $\hat{\boldsymbol{\Sigma}}_G$ by shrinking it towards $\boldsymbol{\Sigma}_P$ can be obtained by substituting the latter for $\boldsymbol{\Omega}$. Adopting an empirical Bayes approach, we replace $\boldsymbol{\Sigma}_P$ with its estimate from a standard, unpenalized (RE)ML analysis, $\hat{\boldsymbol{\Sigma}}_P^0$ (Meyer *et al.* 2011). Letting ν take on the rôle of the tuning factor, gives penalty

$$\mathcal{P}_\Sigma(\boldsymbol{\theta}) \propto C \log |\hat{\boldsymbol{\Sigma}}_G| + \text{tr}(\hat{\boldsymbol{\Sigma}}_G^{-1} \hat{\boldsymbol{\Sigma}}_P^0) \quad \text{with} \quad C = (\psi + q + 1)/\psi \quad (5)$$

If C is approximated with unity, $\mathcal{P}_\Sigma(\boldsymbol{\theta})$ is proportional to the Kullback-Leibler divergence between $\hat{\boldsymbol{\Sigma}}_G$ and $\hat{\boldsymbol{\Sigma}}_P^0$, which is the entropy loss $L_1(\cdot)$ (Eq. 1) with $\boldsymbol{\Sigma}$ and $\hat{\boldsymbol{\Sigma}}$ exchanged (Levina *et al.* 2008).

Based on empirical evidence that estimates of genetic (r_G) and phenotypic (r_P) correlations are often similar, Cheverud (1988) proposed to substitute r_P for r_G if the data did not support accurate estimation of r_G . A more flexible alternative is to penalize the divergence between estimates of the genetic (\mathbf{R}_G) and phenotypic correlation (\mathbf{R}_P) matrix, i.e. to shrink $\hat{\mathbf{R}}_G$ towards $\hat{\mathbf{R}}_P^0$. Analogous to (Eq. 5), this can be achieved using a penalty

$$\mathcal{P}_R(\boldsymbol{\theta}) \propto C \log |\hat{\mathbf{R}}_G| + \text{tr}(\hat{\mathbf{R}}_G^{-1} \hat{\mathbf{R}}_P^0) \quad (6)$$

Similarly, we can use this type of penalty to shrink an estimated covariance matrix towards a chosen structure, akin to the empirical Bayesian approach considered by Chen (1979). For instance, a highly parsimonious description of $\boldsymbol{\Sigma}_G$ can be obtained assuming a factor-analytic structure, fitting a low number of factors. In some cases, we may then want to allow for a data-driven compromise between this structure and an unstructured matrix. A suitable penalty to achieve this with penalized (RE)ML can be obtained by substituting an unpenalized, structured estimate of $\boldsymbol{\Sigma}_G$ for $\hat{\boldsymbol{\Sigma}}_P^0$ in (Eq. 5).

Tuning factors. A crucial question is how to determine the appropriate value of ψ for a given analysis. In a Bayesian vein, this might be chosen *a priori*, analogous to the degree of belief. Hayes and Hill (1981) advocated to base the degree of ‘bending’ on the sample size whilst imposing sufficient shrinkage to ensure $\hat{\Sigma}_G$ was positive definite. Similarly, Meyer (2011) proposed to apply a relatively mild degree of penalization with ψ chosen so that the deviation of $\log \mathcal{L}(\theta)$ from the maximum (at $\psi=0$) was small, arguing that this was likely to exploit some of the benefits of penalized estimation whilst safe-guarding against excessive shrinkage. A natural choice was a limit of $-\frac{1}{2}\chi^2_\alpha$ for one degree of freedom, i.e. the critical value in a likelihood ratio test to detect a significant change in a single parameter at an error probability of α . In a simulation study for 5 traits with $\alpha=0.05$ this yielded reductions in loss for small samples of around 90% of those achieved when exploiting knowledge of the population values to determine ψ .

Most studies concerned with regularization of covariance matrices employ a cross-validation (CV) strategy to estimate the ‘optimal’ value of ψ . This involves splitting the data into so-called training and validation sets. Estimates based on the training data are then obtained for a range of possible values of ψ and corresponding values for a criterion used to assess how well the estimates fit the data – such as $\log \mathcal{L}(\theta)$ – are calculated for the validation set. Typically, this is repeated several times, e.g. in a K -fold CV scheme where each fold in turn is used as validation set with the remainder forming the training set (e.g. Hastie *et al.* 2001). The value of $\hat{\psi}$ is then chosen as that for which the average of the criterion is ‘best’. Clearly, CV is not only a laborious strategy but $\hat{\psi}$ may also be estimated with considerable error which can reduce the efficacy of penalized estimation.

Literature reports on the performance of CV generally pertain to analyses estimating a single covariance matrix only where representative sub-sampling of data sets is straightforward. This is not the case for data with arbitrary genetic relationship structure and fixed effects with potentially small subclasses – which is common for records from livestock improvement schemes. Future work is needed to establish suitable strategies for such scenarios. Additional problems arise with the use of CV in conjunction with penalized (RE)ML: For small samples – and even smaller subsets – the likelihood surface in the vicinity of the maximum tends to be flat, so that the maximum often can not be located accurately. Together with a strong chance of encountering estimates at the boundary of the parameter space, this can lead to ‘validation’ curves which are somewhat jagged or have unexpected jumps. In turn, this can be detrimental to the adequate performance of the CV procedure.

SAMPLING PROPERTIES OF PENALIZED ESTIMATES

An extensive simulation has been carried out to examine the performance of penalized estimation imposing different penalties and employing various strategies to determine the tuning factor. Data were simulated for $q=5$ traits, assumed to be multivariate normally distributed, measured on each of 10 progeny of 100 unrelated sires. A total of 60 sets of population values were considered, with varying levels and spread of heritabilities, genetic and residual correlations and canonical eigenvalues. Details and additional results are given in Meyer *et al.* (2011) and Meyer (2011).

Penalties compared were $\mathcal{P}_\lambda(\theta)$, $\mathcal{P}_\lambda^\ell(\theta)$, $\mathcal{P}_\beta(\theta)$ and $\mathcal{P}_\Sigma(\theta)$. For each, REML estimates of Σ_G and the residual covariance, Σ_E , were obtained for a range of 311 values of ψ from 0 to 1000. Three strategies to determine ψ were employed: 1) Using the known population values to construct matrices of mean squares and cross-products between and within sires, $\hat{\psi}$ was chosen as the value which maximized the unpenalized likelihood $\log \mathcal{L}(\theta)^\psi$, for data represented by these matrices. This can be thought of as sampling an infinite number of additional data sets for the same data structure (strategy V_∞). 2) Using K -fold cross-validation as described above, with $K=3$ (strategy CV3). 3) Finally, $\hat{\psi}$ was chosen as the largest value for which $|\log \mathcal{L}(\theta)^\psi - \log \mathcal{L}(\theta)^0|$, i.e. the reduction in the unpenalized likelihood due to penalization from the maximum (at $\psi=0$) did not exceed $\frac{1}{2}\chi^2_{0.05}$ for 1 degree of freedom, i.e. 1.92 (strategy L5%). A total of 1000 replicates were carried out for each case. The effect

Table 1. Mean PRIAL in estimates of covariance matrices

	Population values				Cross-validation				Likelihood			
	$\mathcal{P}_\lambda(\theta)$	$\mathcal{P}_\lambda^\ell(\theta)$	$\mathcal{P}_\beta(\theta)$	$\mathcal{P}_\Sigma(\theta)$	$\mathcal{P}_\lambda(\theta)$	$\mathcal{P}_\lambda^\ell(\theta)$	$\mathcal{P}_\beta(\theta)$	$\mathcal{P}_\Sigma(\theta)$	$\mathcal{P}_\lambda(\theta)$	$\mathcal{P}_\lambda^\ell(\theta)$	$\mathcal{P}_\beta(\theta)$	$\mathcal{P}_\Sigma(\theta)$
Σ_G	35.8	71.3	68.1	70.6	23.1	55.9	61.2	54.9	41.4	68.3	69.8	64.9
Σ_E	57.9	43.4	59.7	13.3	14.1	26.7	38.0	10.7	43.4	35.0	53.9	12.0
Σ_P	1.1	1.2	1.2	1.2	-0.4	0.4	0.2	0.2	-0.7	0.7	0.4	0.4

of penalization on estimates of covariance matrices was then summarized as *percentage reduction in average loss*, $\text{PRIAL} = 100 [\bar{L}_1(\Sigma_x, \hat{\Sigma}_x^0) - \bar{L}_1(\Sigma_x, \hat{\Sigma}_x^\psi)] / \bar{L}_1(\Sigma_x, \hat{\Sigma}_x^0)$ with $\hat{\Sigma}_x^0$ and $\hat{\Sigma}_x^\psi$ the unpenalized and penalized estimates, respectively, and $\bar{L}_1(\cdot)$ the entropy loss in $\hat{\Sigma}_x$ averaged over replicates. In addition, the relative bias (in %) in estimates of canonical eigenvalues and heritabilities was calculated as $100(\hat{\lambda}_i - \lambda_i) / \lambda_i$ and $100(\hat{h}_i^2 - h_i^2) / h_i^2$, respectively.

Reduction in loss. Table 1 gives the average PRIAL obtained across the 60 cases for the different penalties and methods to determine ψ . Mean values hide considerable variation in ranking of penalties for individual cases. While none was best throughout, penalties on canonical eigenvalues assuming a common mean tended to perform better than $\mathcal{P}_\Sigma(\theta)$ and $\mathcal{P}_\beta(\theta)$ when population values for the λ_i were fairly similar. Conversely, $\mathcal{P}_\Sigma(\theta)$ and $\mathcal{P}_\beta(\theta)$ mostly yielded larger PRIALs for the other cases.

As reported by Meyer and Kirkpatrick (2010), taking logarithms of the canonical eigenvalues ($\mathcal{P}_\lambda^\ell(\theta)$) greatly improved the efficacy of a penalty on the variance among the estimated eigenvalues on estimates of Σ_G . For strategies V_∞ and L5% this was accompanied by a reduction in PRIAL for $\hat{\Sigma}_E$. This could be attributed to cases with population values λ_i spread far apart for which $\mathcal{P}_\lambda(\theta)$ yielded a substantial reduction in loss for $\hat{\Sigma}_E$ but yielded poor results for $\hat{\Sigma}_G$. For strategies V_∞ and CV3, there was little difference in PRIAL for $\hat{\Sigma}_G$ between penalties $\mathcal{P}_\lambda^\ell(\theta)$ and $\mathcal{P}_\Sigma(\theta)$. However, values for $\hat{\Sigma}_E$ for $\mathcal{P}_\Sigma(\theta)$ were substantially lower, as this penalty involved $\hat{\Sigma}_G$ only. Conversely, penalty $\mathcal{P}_\beta(\theta)$ resulted in the highest PRIAL for $\hat{\Sigma}_E$. This can be explained by $\mathcal{P}_\Sigma(\theta)$ penalizing both λ_i and $1 - \lambda_i$, which, for $\Sigma_E = \mathbf{T}(\mathbf{I} - \mathbf{\Lambda})\mathbf{T}'$, is equivalent to a direct penalty on Σ_E as well as Σ_G . Placing a quadratic penalty on both λ_i and $1 - \lambda_i$ yielded comparable results (Meyer 2011). Interestingly, $\mathcal{P}_\beta(\theta)$ was least affected by errors in estimates of ψ for strategies CV3 and L5%.

Bias. Corresponding relative biases in estimates of canonical eigenvalues and heritabilities (h_i^2) obtained using cross-validation to determine ψ are shown in Table 2. As expected from theory, unpenalized estimates of the largest $\hat{\lambda}_i$ were biased upwards and of the smallest $\hat{\lambda}_i$ were biased downwards, with the large value for $\hat{\lambda}_5$ an artifact of small population values. On average, shrinkage of the λ_i towards their mean caused a downwards bias in $\hat{\lambda}_1$. Whilst taking logarithms ($\mathcal{P}_\lambda^\ell(\theta)$) alleviated this bias, it also resulted in a substantial upwards bias in $\hat{\lambda}_5$. However, as the smallest $\hat{\lambda}_i$ contribute least to estimates of Σ_G , the PRIAL for $\mathcal{P}_\lambda^\ell(\theta)$ was substantially higher than for $\mathcal{P}_\lambda(\theta)$. For penalty $\mathcal{P}_\Sigma(\theta)$ bias in the largest $\hat{\lambda}_i$ was very similar to those in unpenalized estimates while the smallest $\hat{\lambda}_i$ were substantially

Table 2. Mean relative bias for CV3

	$\psi=0$	$\mathcal{P}_\lambda(\theta)$	$\mathcal{P}_\lambda^\ell(\theta)$	$\mathcal{P}_\beta(\theta)$	$\mathcal{P}_\Sigma(\theta)$
$\hat{\lambda}_1$	9.5	-11.3	-3.7	-7.8	8.3
$\hat{\lambda}_2$	26.5	15.9	16.7	20.6	25.5
$\hat{\lambda}_3$	16.7	22.0	26.8	26.2	25.3
$\hat{\lambda}_4$	-19.4	10.8	53.3	28.4	42.1
$\hat{\lambda}_5$	-78.8	-25.6	107.0	34.8	86.7
\hat{h}_1^2	-1.1	-14.0	-6.7	-10.8	0.9
\hat{h}_2^2	3.8	-5.1	4.1	-0.1	10.5
\hat{h}_3^2	4.5	-0.5	11.1	5.7	16.2
\hat{h}_4^2	7.2	7.2	23.1	14.8	26.6
\hat{h}_5^2	12.3	19.5	44.7	30.6	45.7

biased upwards, albeit somewhat less than from penalized estimation using $\mathcal{P}_\lambda^\ell(\theta)$.

Population values for h_i^2 declined with trait number. Biases in unpenalized estimates of heritabilities were small, with some effect of constraints on the parameter noticeable which biased h_1^2

downwards and the other values upwards. Penalized estimation increased bias, especially for the extreme values, illustrating the trade-off between sampling variance and bias to reduce loss. Differences between penalties were similar to those observed for the canonical eigenvalues. Results for strategies L5% and V_∞ exhibited a comparable pattern (not shown) with somewhat larger biases for V_∞ .

Similarly, unpenalized estimates of genetic correlations were slightly biased, with a mean deviation from population values of -0.019 and a mean absolute deviation of 0.033 . Corresponding values for strategy V_∞ were -0.030 and 0.064 for $\mathcal{P}_\lambda(\theta)$, -0.046 and 0.101 for $\mathcal{P}_\lambda^\ell(\theta)$, -0.043 and 0.094 for $\mathcal{P}_\beta(\theta)$, and -0.039 and 0.085 for $\mathcal{P}_\Sigma(\theta)$. Again, biases tended to increase with the associated PRIAL, though at comparable PRIALs due to $\mathcal{P}_\lambda^\ell(\theta)$ and $\mathcal{P}_\Sigma(\theta)$, the latter resulted in less bias in estimates of r_G . As for the other quantities examined, differences between penalties became somewhat blurred for strategies to determine ψ which did not rely on knowledge of the population parameters.

APPLICATION: CARCASS TRAITS FOR BEEF CATTLE

Carcass characteristics are a typical example of traits that are ‘hard to measure’ but are of major importance in livestock improvement programmes. Traits considered were carcass weight (WT), eye muscle area (EMA), percentage intra-muscular fat (IMF), retail beef yield (RBY), and fat thickness at the P8 site on the rump (P8) and the 12th/13th rib (RIB) of Hereford cattle. Data were collected at abattoirs as part of a meat quality research project (CRC I) and have been analysed previously; see Reverter *et al.* (2000) for further details. There were 1030 animals in the data, all of which had WT recorded. Numbers of measurements for the other 5 traits were 864 (EMA), 992 (IMF), 370 (RBY), 999 (P8) and 1014 (RIB). All records were pre-adjusted for differences in age at slaughter or carcass weight as described in Reverter *et al.* (2000). Only 30% of individuals had all 6 traits recorded, but 54% had 5 traits measured. Animals in the data were the progeny of 59 sires. Adding pedigree information yielded a total of 2817 animals.

The model of analysis was a simple animal model, fitting animals’ additive genetic effects as random effects. The only fixed effects fitted were ‘contemporary groups’ (CG) which represented a combination of herd of origin, sex of animal, date of slaughter, abattoir, finishing regime and target market subclasses, with up to 180 levels per trait. Estimates of Σ_G and Σ_E were obtained by REML as described in Meyer and Kirkpatrick (2010) using WOMBAT (Meyer 2007a), considering penalties $\mathcal{P}_\lambda^\ell(\theta)$ and $\mathcal{P}_\Sigma(\theta)$, as defined above. Tuning factors ψ were determined using 4 repeats of CV with $K=3$ (CV3) and, for $\mathcal{P}_\Sigma(\theta)$ only, CV with $K=10$ (CV10). To minimize problems due to splitting small CG subclasses, data were subdivided by randomly assigning all animals in a CG (for WT) to a subset. Splits were repeated until all subsets comprised between 29 and 37% and between 8.5 and 11.5% of records for $K=3$ and $K=10$, respectively. Results were contrasted to ψ obtained by limiting the change in $\log \mathcal{L}(\theta)$ to approximately $-\frac{1}{2}\chi^2 0.05$ for 1 degree of freedom (L5%).

Results. Estimates of heritabilities from different analyses (\pm approximate standard errors for $\psi=0$) together with the value for ψ and the resulting change (Δ) in $\log \mathcal{L}(\theta)$ are summarized in Table 3. Using CV3 to estimate ψ suggested a more severe degree of penalization than L5%, especially for penalty $\mathcal{P}_\lambda^\ell(\theta)$. With small numbers of records for individual traits, standard errors for unpenalized estimates were substantial. Different types of penalty

Table 3. Heritability estimates for carcass traits

	No penalty	$\mathcal{P}_\lambda^\ell(\theta)$		$\mathcal{P}_\Sigma(\theta)$		
		L5%	CV3	L5%	CV3	CV10
ψ	0	2.90	9.50	9.50	17.00	9.75
$\Delta \log \mathcal{L}$	0	-1.927	-5.077	-1.914	-3.155	-2.106
WT	0.590 \pm 0.135	0.532	0.482	0.603	0.615	0.604
EMA	0.643 \pm 0.154	0.575	0.464	0.665	0.679	0.665
IMF	0.353 \pm 0.122	0.349	0.347	0.390	0.416	0.391
RBY	0.331 \pm 0.176	0.329	0.340	0.389	0.427	0.390
P8	0.207 \pm 0.093	0.261	0.294	0.285	0.316	0.287
RIB	0.251 \pm 0.095	0.289	0.308	0.305	0.331	0.306

and different strategies to select ψ changed results to varying degrees. However, all penalized estimates were well within the range of the 95% confidence intervals of the unpenalized values. As expected from simulation results (see Table 2), using $\mathcal{P}_\lambda^\ell(\theta)$ decreased estimates of the largest values, while both penalties increased the smallest values similarly. Unpenalized estimates of canonical eigenvalues ranged from 0.76 to 0.04. Imposing a penalty decreased this to 0.66 – 0.14 (L5%) and 0.53 – 0.21 (CV3) for $\mathcal{P}_\lambda^\ell(\theta)$ and 0.76 – 0.14 (L5%) and 0.76 – 0.18 (CV3) for $\mathcal{P}_\Sigma(\theta)$.

Corresponding estimates of genetic correlations are contrasted in Figure 1. Shown for each pair of traits are the unpenalized estimate together with the range given by plus and minus one standard deviation, flanked by estimates obtained using $\mathcal{P}_\lambda^\ell(\theta)$ (left side) and $\mathcal{P}_\Sigma(\theta)$ (right side), selecting ψ using strategies L5% and CV3. For most correlations, penalized estimation reduced the magnitude (sign ignored) compared to unpenalized values. However, changes were relatively small, with average values of –0.06 (L5%) and –0.12 (CV3) for $\mathcal{P}_\lambda^\ell(\theta)$ and –0.06 (L5%) and –0.07 (CV3) for $\mathcal{P}_\Sigma(\theta)$. With the exception of correlations between EMA or P8 with RIB, average changes in estimates (over the different penalties applied) were markedly less than one standard deviation. Other studies have generally reported little genetic association between EMA and RIB, either for carcass traits or corresponding measures obtained via live ultrasound scanning (e.g. Meyer 2007b). Hence, the unpenalized estimate of 0.69 ± 0.18 in these data appeared too high and the reduction to 0.5 or less (0.33 for $\mathcal{P}_\lambda^\ell(\theta)$ with $\psi=9.5$) is plausible. In contrast, an estimate of 0.80 ± 0.17 for P8 and RIB agreed well with literature results. Presumably the consistent, relatively large change in this parameter due to penalization was, to some extent at least, an artifact of the change in r_G^2 between EMA and RIB.

DISCUSSION

We have outlined an extension of current, standard methodology to estimate genetic parameters in a mixed model framework that has the scope to yield ‘better’ estimates, especially for multivariate analyses comprising more than just a few traits. This is achieved by penalizing the likelihood, with the penalty a function of the parameters aimed at reducing sampling variation. A number of suitable penalties have been described with emphasis on those ‘borrowing strength’ from estimates of the phenotypic covariance or correlation matrices which are typically estimated much more accurately than their genetic counterparts.

All penalties presented have a Bayesian motivation, i.e. can be derived assuming certain prior distributions for covariance matrices or their eigenvalues. In contrast to full Bayesian analyses, location or scale parameters for the priors are estimated from the data at hand, i.e. our penalized maximum likelihood procedure can be considered as analogous to an empirical Bayes approach.

Simulation results have been presented, both here and in companion papers (Meyer *et al.* 2011; Meyer 2011), demonstrating that substantial reductions in loss, i.e. the difference between true and estimated values, can be achieved for estimates of genetic covariance matrices. As expected, this comes at the price of increasing bias, over and above that introduced by constraining estimates to the parameter space in standard analyses. The magnitude and direction of the additional bias depend on the population parameters and penalty applied, but in general penalization caused estimates of the highest heritabilities to be reduced and those of the smallest heritabilities to be increased while estimates of genetic correlations were reduced in absolute value. As illustrated in the applied example,

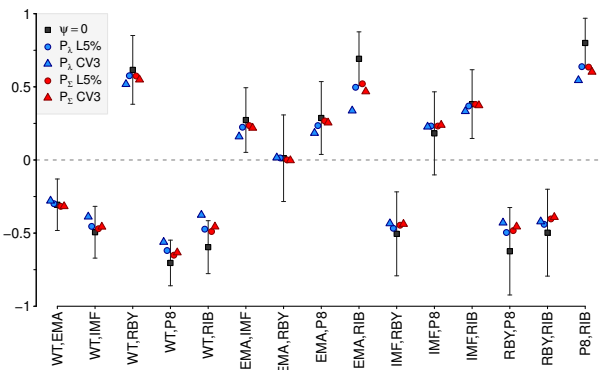


Figure 1. Estimates of genetic correlations.

for small samples these changes were usually well within the confidence intervals of the unpenalized estimates. With comparable reductions in loss to other penalties, $\mathcal{P}_R(\theta)$ which shrinks the genetic towards the phenotypic correlation matrix appeared to result in least bias (Meyer *et al.* 2011).

The underlying motivation for the use of penalized estimation, of course, is the belief that improved estimates of genetic parameters directly translate into better predictions of animals' genetic merit and more appropriate selection decisions, in particular when weighing information on different traits. Hayes and Hill (1981) demonstrated that use of 'bending' substantially improved the achieved response to index selection. Further work should examine the effectiveness of the methodology and new penalties presented in such context.

CONCLUSIONS

Penalized maximum likelihood estimation provides the means to 'make the most' of limited and precious data and facilitates more stable estimation for multi-dimensional analyses even when samples are somewhat larger. We anticipate that it will become part of our everyday toolkit as truly multivariate estimation for quantitative genetic problems becomes routine.

ACKNOWLEDGEMENTS

This work was funded by Meat and Livestock Australia. A reviewer of Meyer and Kirkpatrick (2010) suggested use of the distribution of order statistics.

REFERENCES

- Chen C.F. (1979) *J. Roy. Stat. Soc. B* **41**:235.
Cheverud J.M. (1988) *Evolution* **42**:958.
Green P.J. (1998) In 'Encyclopedia of Statistical Sciences', Vol. 2. Wiley, pp. 578–586.
Hastie T., Tibshirani R. and Friedman J. (2001) 'The Elements of Statistical Learning'. Springer.
Hayes J.F. and Hill W.G. (1981) *Biometrics* **37**:483.
Hoerl A.E. and Kennard R.W. (1970) *Technometrics* **12**:69.
James W. and Stein C. (1961) In 'Fourth Berkeley Symposium on Mathematical Statistics and Probability', Vol. 1. pp. 361–379.
Lawley D.N. (1956) *Biometrika* **43**:128.
Levina E., Rothman A.J. and Zhu J. (2008) *Ann. Appl. Stat.* **2**:245.
Meyer K. (2007a) *J. Zhejiang Uni. SCIENCE B* **8**:815.
Meyer K. (2007b) *J. Anim. Breed. Genet.* **124**:50.
Meyer K. (2011) *Proc. Ass. Advan. Anim. Breed. Genet.* **19**:83.
Meyer K. and Kirkpatrick M. (2010) *Genetics* **185**:1097.
Meyer K., Kirkpatrick M. and Gianola D. (2011) *Proc. Ass. Advan. Anim. Breed. Genet.* **19**:87.
Reverter A., Johnston D.J., Graser H.U., Wolcott M.L. and Upton W.H. (2000) *J. Anim. Sci.* **78**:1786.
Sorensen D. and Gianola D. (2002) 'Likelihood, Bayesian and MCMC Methods in Quantitative Genetics'. Springer.
Stein C. (1975) In 'Reitz lecture'. 39th Annual Meeting Institute Mathematical Statistics. Atlanta.
Thompson R., Brotherstone S. and White I.M.S. (2005) *Phil. Trans. R. Soc. B* **360**:1469.
Tibshirani R. (1996) *J. Roy. Stat. Soc. B* **58**:267.
Tyrisevä A.M., Meyer K., Fikse F., Ducrocq V., Jakobsen J., Lidauer M.H. and Mäntysaari E.A. (2011) *Genet. Select. Evol.* **43**:21.
Warton D.I. (2008) *J. Amer. Stat. Ass.* **103**:340.
Witten D.M. and Tibshirani R. (2009) *J. Roy. Stat. Soc. B* **71**:615.
Yap J.S., Fan J. and Wu R. (2009) *Biometrics* **65**:1068.