

# Tuning-free penalized maximum likelihood to improve estimates of genetic covariance matrices

Karin Meyer

Animal Genetics and Breeding Unit<sup>1</sup>, University of New England, Armidale NSW 2351,  
Australia

## 1 Introduction

Partitioning of the phenotypic variation into its causal components to determine genetic parameters is one of the basic tasks in quantitative genetics. For multiple characteristics of interest, this involves the estimation of covariance matrices due to genetic and other effects. It is well known that such estimates can be subject to substantial sampling variation, as the number of parameters to be estimated generally increases quadratically with the number of traits considered. In particular, it has been recognized early on that the eigenvalues of sample covariance matrices are systematically biased – with the largest eigenvalues overestimated and the smallest values underestimated – and that a major proportion of the sampling variation of covariance matrices can be attributed to this over-dispersion (Lawley, 1956; Ledoit & Wolf, 2004).

A measure of the quality of an estimator is its loss, i.e. the deviation of the estimate from the true value. This is an aggregate of bias and sampling variation. We speak of improving an estimator if we can modify it so that the (expected) loss is lessened. Typically, this involves a reduction in variance at the expense of some additional bias. Work on ‘regularized’ estimation of covariance matrices dates back to the 1970s and earlier, in particular James & Stein (1961) and Stein (1975). Recently, there has been a resurgence in interest especially for higher dimensional matrices, e.g. Bickel & Levina (2008), Warton (2008), Witten & Tibshirani (2009), Ye & Wang (2009), Rothman *et al.* (2010), Bien & Tibshirani (2011), Fisher & Sun (2011), Ledoit & Wolf (2012) or Won *et al.* (2013).

While the statistical literature by and large considers estimation of a single covariance matrix, quantitative genetic analyses involve at least two such matrices, namely due to additive genetic and residual effects. In addition, partitioning of the total variation into its components creates substantial sampling correlations and tends to exacerbate the effects of sampling variation. Nevertheless, there has been comparatively little interest in regularization of multivariate estimates. In a classic paper, Hayes & Hill (1981) proposed to modify estimates of the genetic covariance matrix ( $\Sigma_G$ ) by shrinking the canonical eigenvalues of  $\Sigma_G$  and the phenotypic covariance matrix ( $\Sigma_P$ ) towards their mean, a procedure they described as ‘bending’ the estimate of  $\Sigma_G$  towards that of  $\Sigma_P$ . The underlying rationale was that  $\Sigma_P$ ,

<sup>1</sup>a joint venture with the NSW Department of Primary Industries

the sum of all causal components, is typically estimated much more accurately than any of its components, so that this would ‘borrow strength’ from the estimate of the phenotypic covariance matrix. The authors showed by simulation that using estimates modified in such manner in constructing selection indices could increase the achieved response to selection markedly, as these were closer to the population values than unmodified estimates. However, no method to determine the optimal amount of shrinkage was given and ‘bending’ was thus primarily used only to modify non-positive definite matrices, and all but forgotten when restricted maximum likelihood (REML) estimation, which allows estimates to be constrained to the parameter space, became a common procedure.

More recently, Meyer & Kirkpatrick (2010) demonstrated that REML estimation subject to a penalty on the likelihood proportional to the variance among canonical eigenvalues yields a procedure analogous to ‘bending’ which can substantially reduce loss in estimates of genetic covariance matrices. Penalized maximum likelihood estimation has a direct link to Bayesian estimation: for a given prior distribution of the parameters of interest, an appropriate penalty can be obtained as minus the logarithmic value of the probability density function. For instance, shrinkage of eigenvalues towards their mean then is equivalent to assuming a Normal distribution and assuming a double exponential prior distribution gives a LASSO type penalty (Huang *et al.*, 2006). Using this principle other, effective penalties to reduce loss in estimates of  $\Sigma_G$  can be obtained invoking a Beta distribution for the canonical eigenvalues or assuming an Inverse Wishart distribution for  $\Sigma_G$  (Meyer *et al.*, 2011; Meyer, 2011).

Along with numerous other studies, Meyer (2011) regulated the stringency of penalization by multiplying the penalty with a so-called ‘tuning factor’,  $\psi$ . This can be estimated from the data at hand by  $k$ -fold cross-validation, as the factor maximising the average, unpenalized likelihood in the validation sets. However, this procedure is laborious and can be afflicted by problems in locating the maximum of a fairly flat likelihood surface for analyses involving many traits and not so large data sets. Hence, a mild degree of penalization was proposed as a pragmatic alternative, with  $\psi$  determined as the largest value for which the resulting decrease in the unpenalized likelihood from its maximum at  $\psi = 0$  ( $\Delta\mathcal{L}$ ) did not exceed a chosen, small value. Simulation results showed this strategy to perform well, yielding estimates with less loss than obtained using cross-validation to estimate  $\psi$  in most cases (Meyer, 2011). A suitable limit for  $\Delta\mathcal{L}$  for the smaller data sets and moderate numbers of traits examined appeared to be the value corresponding to the significance level in a likelihood ratio test for a single parameter. While not having to justify a significant change in likelihood due to penalization is intuitively appealing, this may not be a useful choice for other scenarios.

In a Bayesian framework, the influence of the prior and thus the amount of regularization is generally specified through the so-called hyperparameters of the distribution which determine its shape, scale or location. This suggests that an alternative, tuning factor-free formulation for the penalty on the likelihood can be obtained by expressing it terms of the distribution-specific (hyper)parameters which are not dependent on the covariance matrices.

For instance, when assuming a Normal prior for the canonical eigenvalues, the regulating parameter is the variance, with the more shrinkage induced the lower its value. While the penalized likelihood involving a tuning factor is, by definition, highest for  $\psi = 0$ , such formulation may then facilitate direct estimation of the regulating parameter, denoted henceforth as  $\nu$ , simultaneously with the covariance components to be estimated.

Another option, not considered so far in this context, is to employ penalties derived assuming non- or weakly informative priors, which do not involve hyperparameters or set default values for them. This paper investigates the properties of penalized REML estimation of genetic covariance matrices using both these ‘tuning-free’ approaches.

## 2 Penalized maximum likelihood estimation

Consider a simple ‘animal model’ for  $q$  traits,  $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{g} + \mathbf{e}$  with  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{g}$  and  $\mathbf{e}$  the vectors of observations, fixed effects, additive genetic and residual effects, respectively, and  $\mathbf{X}$  and  $\mathbf{Z}$  the corresponding incidence matrices. Let  $\Sigma_G$  and  $\Sigma_E$  represent the matrices of additive genetic and residual covariances among the  $q$  traits, with  $\boldsymbol{\theta}$ , of length  $q(q+1)$ , the vector of parameters to be estimated, comprised of the distinct elements of  $\Sigma_G$  and  $\Sigma_E$ . Further, let  $\text{Var}(\mathbf{y}) = \mathbf{V}$  denote the covariance matrix of the vector of observations. Omitting a constant, the pertaining REML log likelihood is

$$\log \mathcal{L}(\boldsymbol{\theta}) \propto -\frac{1}{2} \left( \log |\mathbf{V}| + \log |\mathbf{X}'_0 \mathbf{V}^{-1} \mathbf{X}_0| + (\mathbf{y} - \mathbf{X}\mathbf{b})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\mathbf{b}) \right) \quad (1)$$

for  $\mathbf{X}_0$  a full-rank submatrix of  $\mathbf{X}$  (e.g. Harville, 1977). Regularized estimates can be obtained by maximizing the *penalized* likelihood

$$\log \mathcal{L}_p(\boldsymbol{\theta}) = \log \mathcal{L}(\boldsymbol{\theta}) - \mathcal{P}(\boldsymbol{\theta}) \quad (2)$$

with the penalty  $\mathcal{P}(\boldsymbol{\theta})$  a selected function of the parameters, aimed at reducing loss in their estimates.

### 2.1 Penalties based on informative priors

As outlined above, borrowing from Bayesian estimation and assuming specific prior distributions for the parameters to be estimated or functions thereof, suitable penalties can be obtained from the probability density function of the priors. We consider some of the cases examined previously (Meyer, 2011).

### 2.1.1 Normal distribution of canonical eigenvalues

Let  $\lambda_i$ , for  $i = 1, q$ , denote the canonical eigenvalues of  $\Sigma_G$  and  $\Sigma_P$ , i.e. the eigenvalues of  $\Sigma_P^{-1}\Sigma_G$  for  $\Sigma_P = \Sigma_G + \Sigma_E$ . A common assumption is that the eigenvalues of a matrix are independently and normally distributed with variance  $\sigma^2$ . This yields a penalty comprised of the quadratic deviation of the eigenvalues from their mean which causes linear shrinkage towards the mean. Assuming a Normal distribution for the  $\lambda_i$  with mean equal to  $\bar{\lambda}$ , the density function is

$$p(\lambda_i) = (2\pi\sigma^2)^{-1/2} \exp\left(-(\lambda_i - \bar{\lambda})^2 / (2\sigma^2)\right) \quad (3)$$

and, taking minus the logarithm of the density, the corresponding penalty for  $\nu = \sigma^2$  is

$$\mathcal{P}_\lambda = \frac{1}{2} \left[ q(\log(\nu) + \log(2\pi)) + \frac{1}{\nu} \sum_{i=1}^q (\lambda_i - \bar{\lambda})^2 \right] \quad (4)$$

### 2.1.2 Beta distribution of canonical eigenvalues

The canonical eigenvalues fall into the interval  $[0, 1]$ . Hence, an obvious alternative is to assume a Beta distribution. This has two scale parameters,  $\alpha$  and  $\beta$ , and density function

$$p(\lambda_i) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \lambda_i^{\alpha-1} (1 - \lambda_i)^{\beta-1} \quad (5)$$

with  $\Gamma(\cdot)$  denoting the Gamma function. The mean and variance of the Beta distribution are  $\alpha/(\alpha + \beta)$  and  $\alpha\beta(\alpha + \beta)^{-2}(\alpha + \beta + 1)^{-1}$ , respectively. This gives matching moments estimators (Evans *et al.*, 2000; chapter 5)

$$\alpha = \bar{\lambda} \nu \quad \text{and} \quad \beta = (1 - \bar{\lambda}) \nu \quad \text{for} \quad \nu = \left[ \bar{\lambda}(1 - \bar{\lambda})(q - 1) / \sum_{i=1}^q (\lambda_i - \bar{\lambda})^2 \right] - 1 \quad (6)$$

This implies that we can specify the Beta prior for a given mean  $\bar{\lambda}$  selecting the single parameter  $\nu$ , also referred to in Bayesian estimation as the prior ‘effective sample size’. The resulting penalty is

$$\begin{aligned} \mathcal{P}_\beta = q \left[ \log \Gamma(\nu) - \log \Gamma(\bar{\lambda}\nu) - \log \Gamma((1 - \bar{\lambda})\nu) \right] &+ (\bar{\lambda}\nu - 1) \sum_{i=1}^q \log(\lambda_i) \\ &+ ((1 - \bar{\lambda})\nu - 1) \sum_{i=1}^q \log(1 - \lambda_i) \end{aligned} \quad (7)$$

### 2.1.3 Inverse Wishart distribution of covariance matrices

A conjugate and thus widely used prior in Bayesian estimation of covariance matrices is the Inverse Wishart (IW) distribution, characterized by a positive definite scale matrix  $\Omega$  and

the degrees of freedom  $\nu$ . For covariance matrix  $\Sigma$ , this has probability density function

$$p(\Sigma) = \left(2^{\nu q/2} \Gamma_q(\nu/2)\right)^{-1} |\Omega|^{\nu/2} |\Sigma|^{-(\nu+q+1)/2} \exp\left(-\text{tr}(\Sigma^{-1}\Omega)/2\right) \quad (8)$$

with  $\Gamma_q(\cdot)$  the multivariate Gamma function. This gives penalty

$$\mathcal{P}_\Sigma = \frac{1}{2} \left[ \nu q \log(2) + (\nu + q + 1) \log |\Sigma| - \nu \log |\Omega| + \text{tr}(\Sigma^{-1}\Omega) \right] \\ + q(q-1)/4 \log(\pi) + \sum_{i=1}^q \log \Gamma((\nu + 1 - i)/2) \quad (9)$$

As the argument of the (univariate) Gamma function needs to be positive, the degrees of freedom are restrained to  $\nu > q - 1$ . This penalty shrinks the estimate of  $\Sigma$  towards  $\Omega$ . Following the idea of borrowing strength from the phenotypic covariance matrix, Meyer *et al.* (2011) proposed to use its estimate as shrinkage target.

The IW prior is often criticized for its lack of flexibility, as the same degrees of freedom are assumed to be appropriate for all  $q(q+1)/2$  distinct elements of  $\Sigma$ . Attempts to alleviate this problem included extensions to a scaled IW (Gelman & Hill, 2007; section 13.3) or a generalized IW prior (Brown, 2006). Penalties based on the latter have been shown to be beneficial for scenarios with greatly differing amounts of data for different traits, but the reduction in loss over and above that achievable using a penalty based on the standard IW prior required substantial additional effort to determine multiple tuning parameters (Meyer, 2013). Other authors considered non-conjugate priors, e.g. a multivariate Normal distribution for the elements of the matrix logarithm (Leonard & Hsu, 1992; Hsu *et al.*, 2010), or decomposed the covariance matrix into correlations and variances and suggested different priors on the two classes of parameters (Daniels & Kass, 1999; Barnard *et al.*, 2000; Gaskins *et al.*, 2013); see also Bouriga & Féron (2013) for a recent review.

## 2.2 Penalties derived from ‘non-informative’ priors

At the other end of the spectrum, less informative priors for Bayesian estimation of covariance matrices have been proposed. Typically, such priors are directed at Bayesian inference when little is known about the parameters of interest, or if there is concern that the prior may overwhelm the data. For covariance matrices, such priors usually generate marginal distributions for individual parameters which are peaked (i.e. somewhat informative) due to requirements for the matrix to be positive definite.

### 2.2.1 ‘Objective’ priors

A number of non-informative priors for covariance matrices have been proposed early on, see Sun & Berger (2007) for review of what they term ‘objective’ priors. Typically, these do not

involve any hyperparameters regulating the strength of shrinkage, with most comprised of a function of the eigenvalues or the determinant of the covariance matrix and its dimension. Yang & Berger (1994) contrasted Jeffreys (independence) prior,

$$p_J(\Sigma) \propto |\Sigma|^{-(q+1)/2} \quad (10)$$

with a reference prior,  $p_R(\Sigma) \propto (|\Sigma| \prod_{i < j} (\gamma_i - \gamma_j))^{-1}$  (with  $\gamma_i$  the eigenvalues of  $\Sigma$  in descending order), emphasizing that the latter tended to put more mass in the region of equal eigenvalues, and thus yielded better eigenstructure shrinkage. Sun & Berger (2007) showed that for  $\Sigma^{-1} = \mathbf{L}'\mathbf{L}$ , with  $\mathbf{L}$  a lower triangular matrix comprising elements  $l_{ij}$ , several such priors suggested in the literature can be expressed in the form  $\prod_{i=1}^q l_{ii}^{-\delta_i}$ , yielding a penalty

$$\mathcal{P}_O = \sum_{i=1}^q \delta_i \log(l_{ii}) \quad (11)$$

As  $\mathbf{L}$  is not unique, neither is the prior or penalty. Values considered included  $\delta_i = 1$  for a one-at-a-time reference prior,  $\delta_i = i$  and  $\delta_i = q - i + 1$ . For instance, for  $\mathbf{L}$  the inverse of the Cholesky factor of  $\Sigma$  and  $\delta_i = 1$ , the penalty based on the one-at-a-time reference prior is  $\mathcal{P}_{R1} = -\frac{1}{2} \log |\Sigma|$ .

### 2.2.2 Weakly informative priors

An alternative gaining popularity are weakly informative priors, i.e. priors set up intentionally so that the information provided is weaker than the actual prior knowledge available (Gelman, 2006).

Huang & Wand (2013) considered a hierarchical specification, with  $\Sigma$  assumed to have an IW distribution with diagonal scale matrix,  $\Omega = 2(\nu - q + 1) \text{Diag}\{\tau_i^{-1}\}$  and the reciprocals of its diagonal elements assumed to have independent Inverse Gamma (IG) distributions,  $\tau_i \sim IG\left(\frac{1}{2}, T_i^{-2}\right)$ , with  $\nu > q - 1$  and  $T_i > 0$ . The authors showed that this yielded a matrix generalization of the half- $t$  prior proposed by Gelman (2006), i.e. that a large values for  $T_i$  lead to an arbitrarily weakly informative prior for the corresponding standard deviation, and that it induced marginal uniform prior distributions for all correlations for  $\nu = q + 1$ . Furthermore, they demonstrated by simulation that their proposed prior yielded Bayesian estimates of sparse covariance matrices with smaller error than the classical IW prior. Omitting some constants, Huang & Wand (2013) gave the resulting density function for  $\Sigma$  as

$$p(\Sigma) \propto |\Sigma|^{-(\nu+q+1)/2} \prod_{i=1}^q \left[ (\nu - q + 1)(\Sigma^{-1})_{ii} + A_i^{-2} \right]^{-(\nu+1)/2} \quad (12)$$

with  $(\Sigma^{-1})_{ii}$  the  $i$ -th diagonal element of  $\Sigma^{-1}$ . This gives penalty

$$\mathcal{P}_{HW} = \frac{\nu + q + 1}{2} \log |\Sigma| + \frac{\nu + 1}{2} \sum_{i=1}^q \log [(\nu - q + 1)(\Sigma^{-1})_{ii} + A_i^{-2}] \quad (13)$$

Motivated by the need to avoid degenerate estimates of covariance matrices in penalized maximum likelihood estimation, Chung *et al.* (2013) proposed a Wishart distribution, with scale matrix a large multiple of the identity matrix, together with a default choice of hyperparameters as a weakly informative prior. For  $\Omega = \frac{1}{2}\tau^{-1}\mathbf{I}$  and  $|\Sigma| = \prod_i \gamma_i$  and  $\exp(-\tau \text{tr}(\Sigma)) = \prod_i \exp(-\tau\gamma_i)$ , this gives prior density

$$p(\Sigma) \propto \tau^{\nu q/2} \left( \Gamma_q(\nu/2) \right)^{-1} \prod_{i=1}^q \gamma_i^{(\nu-q-1)/2} \exp(-\tau\gamma_i) \quad (14)$$

(Chung *et al.*, 2013) and corresponding penalty

$$\mathcal{P}_C = \log \left( \Gamma_q \left( \frac{\nu}{2} \right) \right) - \frac{\nu q}{2} \log \tau + \tau \sum_{i=1}^q \gamma_i - \frac{\nu - q - 1}{2} \sum_{i=1}^q \log(\gamma_i) \quad (15)$$

The authors suggested default choices of  $\nu = q + 2$  and a value of  $\tau$  tending towards zero (using  $\tau = 10^{-4}$  in their applied example). For  $\nu = q + 2$  and  $\tau$  sufficiently small, this implies a penalty dominated by  $-\frac{1}{2} \log |\Sigma|$ , similar to  $\mathcal{P}_O$  for  $\delta_i = 1$ . Chung *et al.* (2013) outlined that this Wishart density implies independent Gamma distributions,  $\Gamma(1.5, \tau)$ , for the eigenvalues of  $\Sigma$ , which have zero mass at the boundary, and thus ensure non-degenerate estimates.

### 3 Simulation study

A simulation study, similar to that used by Meyer (2011) to evaluate penalized estimation involving a tuning factor, was carried out to examine the efficacy of the penalties proposed above.

#### 3.1 Data

Data were sampled from multivariate normal distributions for  $q = 5$  (for comparison with previous analyses only) and  $q = 9$  traits, assuming a balanced paternal half-sib design comprised of  $s$  unrelated sires families of size  $n=10$  each. Sample sizes considered were  $s = 100, 400$  and  $1000$ . A wide range of population values for genetic and residual variance components were chosen, obtained by combining 12 constellations of heritabilities (A to L for  $q = 5$  and M to X for  $q = 9$ ) with different correlation structures ( $I$  to  $\mathcal{V}$  and  $I, \mathcal{V}I$  to  $IX$  for  $q = 5$  and  $q = 9$ , respectively) to generate 60 sets for each number of traits, labelled A-I to X-IX in the following. Details are summarized in Appendix A. A total of 500 samples per case and sample size were acquired.



## 3.2 Analyses

REML estimates of  $\Sigma_G$  and  $\Sigma_E$  for each sample were obtained for different types of penalties as well as penalty-free, using a Method of Scoring algorithm together with simple derivative-free search steps to locate the maximum of the (penalized) likelihood function. This was done using a parameterisation to the elements of the canonical decomposition, as described by Meyer and Kirkpatrick Meyer & Kirkpatrick (2010), restraining estimates of  $\lambda_i$  to the interval of  $[0.0001, 0.9999]$ .

Regulating parameters  $\nu$  were formulated as a function of a variable  $\kappa$ , a selected minimum value  $\mu$ , and a function of the number of traits depending on the type of penalty,  $f(q)$ ,

$$\nu = \mu + \exp(\kappa) + f(q) \quad (16)$$

Unless otherwise specified,  $\mu$  was set to 0.01. Analyses were performed for selected, fixed values of  $\kappa$  and attempting to estimate the optimal value for each sample directly. Direct estimation (DE) was carried out by evaluating points on the profile likelihood for  $\kappa$ , i.e. maximizing  $\log \mathcal{L}_P(\theta)$  with respect to the covariance components to be estimated for chosen values of  $\kappa$ , determining the value for  $\kappa$  (and thus  $\nu$ ) maximizing  $\log \mathcal{L}_P(\theta)$  through a quadratic approximation of the profile. This was done using Powell's (2006) FORTRAN subroutine NEWUOA. The formulation chosen ensured a minimum value for  $\nu$  of  $\mu + f(q)$ . An upper limit of  $20 + \mu + f(q)$ , decided upon after initial investigations to establish a sensible limit, was enforced by setting  $-\log \mathcal{L}_P(\theta)$  to a large value of  $10^{20}$  if  $\kappa$  exceeded  $\log(20)$ .

## 3.3 Penalties

A total of 12 different penalties, summarised in Table 1 were examined. There were 5 penalties targetting the canonical eigenvalues, similar to those considered by Meyer (2011):  $\mathcal{P}_\lambda$  provided the equivalent to 'bending' of  $\Sigma_G$ , shrinking the canonical eigenvalues,  $\lambda_i$ , towards their mean.  $\mathcal{P}_{l\lambda}$  applied the same penalty to the eigenvalues transformed to logarithmic scale, denoted by the subscript  $l\lambda$ , i.e. assumed a log-Normal distribution as suggested by Daniels & Kass (2001). The canonical eigenvalues of  $\Sigma_P^{-1}\Sigma_E$  are  $1 - \lambda_i$ . Hence, imposing the same penalty on  $\Sigma_E$  in addition to  $\Sigma_G$  is equivalent to  $\mathcal{P}_\lambda$  multiplied by a factor of 2 (original scale). On the logarithmic scale, this is different and  $\mathcal{P}_{l\lambda}^+$  and  $\mathcal{P}_{l\lambda}^*$  summed penalties given by (Eq. 4) applied to  $\log(\lambda_i)$  and  $\log(1 - \lambda_i)$ , assuming the same and allowing for different regulation parameters, respectively (using the subscript '\*' to flag analyses requiring estimation of two parameters). Both were examined as calculated and taking their average ( $s = 2$ ). In addition, analyses were carried out for fixed values of  $\nu = 0.04$  to  $0.08$  for  $\mathcal{P}_\lambda$  and  $\nu = 0.5$  to  $2$  for  $\mathcal{P}_{l\lambda}^+$ .

Likewise, a penalty based on the Beta distribution,  $\mathcal{P}_\beta$ , involved multiples of  $\log(\lambda_i)$  and  $\log(1 - \lambda_i)$ . The probability density function of this distribution if one or both of its scale parameters are less than unity is J- or U-shaped, i.e. can have substantial mass close to



the extremes which is undesirable. Hence, values of  $\alpha = 1 + \bar{\lambda}\nu$  and  $\beta = 1 + (1 - \bar{\lambda})\nu$  were used in setting the penalty for a given effective size  $\nu$ . Whilst this changed the mean of the assumed prior distribution to  $(\bar{\lambda}\nu + 1)/(\nu + 2)$  (and corresponding mode to  $\bar{\lambda}$ ), it ensured a peaked distribution. An alternative which maintained the mean of  $\bar{\lambda}$  would have been to set  $\beta = \alpha(1 - \bar{\lambda})/\bar{\lambda}$  for  $\alpha = 1 + \bar{\lambda}\nu$  or  $\alpha = \beta\bar{\lambda}/(1 - \bar{\lambda})$  for  $\beta = 1 + (1 - \bar{\lambda})\nu$ . However, preliminary investigation showed this to perform less well overall, and this option was thus not pursued. For direct estimation, a minimum value for  $\kappa$  of  $\mu = 1$  was used. Further analyses for  $\mathcal{P}_\beta$  considered fixed values for  $\kappa$  ranging from 1 to 20.

Three penalties derived from the IW distributions were considered.  $\mathcal{P}_\Sigma$  imposed a penalty for  $\Sigma_G$  only,  $\mathcal{P}_\Sigma^+$  and  $\mathcal{P}_{\Sigma^*}^+$  involved the sum of penalties on  $\Sigma_G$  and  $\Sigma_E$  assuming the same and different regulation parameters, respectively. As given in (Eq. 9), the resulting degree of penalization was rather severe in most cases, even for the lowest possible degrees of freedom. Hence, penalties were scaled by a factor of  $s^{-1}$  to reduce their impact, considering values of  $s = 2, \sqrt{q}, (q + 1)/2, q$  and  $2q$ . Furthermore, fixed values of  $\kappa$  ranging from 0.1 to 9 were examined.

For comparison, four additional penalties based on non- or weakly informative priors were considered in addition. All were applied to both  $\Sigma_G$  and  $\Sigma_E$ , as initial investigations had shown that penalization of  $\Sigma_G$  only was too detrimental for estimates of  $\Sigma_E$  and thus  $\Sigma_P$ . Derived from the Jeffreys' rule prior and the one-at-a-time reference prior (Sun & Berger (2007); see (Eq. 11) for  $\delta = 1$ ),  $\mathcal{P}_J$  and  $\mathcal{P}_{R1}$  did not involve a regulating parameter but were simply proportional to the matrix determinant on the logarithmic scale

$$\mathcal{P}_J = \frac{q + 1}{2} \log |\Sigma| \quad (17)$$

$$\mathcal{P}_{R1} = \frac{1}{2} \log |\Sigma| \quad (18)$$

Penalties  $\mathcal{P}_{HW}$  and  $\mathcal{P}_{Ch}$ , given in (Eq. 13) and (Eq. 15), respectively, used values setting the scale matrices as suggested by Huang & Wand (2013) and Chung *et al.* (2013),  $A = 10^5$  for  $\mathcal{P}_{HW}$  and  $\tau = 10^{-4}$  for  $\mathcal{P}_{Ch}$ . Both involved a regulating parameter, with default values of  $\nu = q + 1$  ( $\mathcal{P}_{HW}$ ) and  $\nu = q + 2$  ( $\mathcal{P}_{Ch}$ ) proposed by the authors for weak informativity. In addition, values of  $q + 2$  to  $q + 8$  were considered. As for the penalties based on the IW distribution with  $\Sigma_P$  as scale matrix,  $\mathcal{P}_J$  and  $\mathcal{P}_{HW}$  yielded very stringent penalties and were thus again scaled by multiplying with  $s^{-1}$ , for  $s$  as described above.

### 3.4 Summary statistics

For each sample and analysis, the quality of estimates was evaluated through their entropy loss (James & Stein, 1961)

$$L_1(\Sigma_X, \hat{\Sigma}_X) = \text{tr}(\Sigma_X^{-1} \hat{\Sigma}_X) - \log |\Sigma_X^{-1} \hat{\Sigma}_X| - q$$

for  $X = G, E$  and  $P$ , with  $\Sigma_X$  denoting the matrix of population values and  $\hat{\Sigma}_X$  the corresponding estimate. As suggested by Lin & Perlman (1985), the percentage reduction in average loss (PRIAL)

$$100 \left[ 1 - \bar{L}_1(\Sigma_X, \hat{\Sigma}_X^\nu) / \bar{L}_1(\Sigma_X, \hat{\Sigma}_X^0) \right]$$

was then used as the criterion to summarize the effects of penalization, with  $\bar{L}_1(\cdot)$  the entropy loss averaged over replicates, and  $\hat{\Sigma}_X^\nu$  and  $\hat{\Sigma}_X^0$  representing the penalized and corresponding unpenalized REML estimate of  $\Sigma_X$ , respectively. In addition, the average reduction in  $\log \mathcal{L}(\theta)$  (the unpenalized likelihood) due to penalization from its maximum (for unpenalized estimates),  $\Delta \mathcal{L}$ , was calculated.

## 4 Results

As reported for penalized estimation with a tuning factor (Meyer, 2011), there were considerable differences in the efficacy of penalized estimation between sets of population values, with no type of penalty ‘best’ for all scenarios. Mean summary statistics across the 60 sets of population values for  $q = 9$  traits are given in Table 2, and corresponding results for individual cases for selected penalties and a medium sample size are shown in Figure 1.

### 4.1 Direct estimation of regulation parameters

Assuming a Normal distribution for the canonical eigenvalues, direct estimation of the regulating variance performed well in most cases, with sufficient – though generally small – curvature in the profile likelihood for  $\nu$  for the quadratic approximation procedure to yield a suitable estimate for the strength of penalization. As to be expected, on the original scale PRIAL was highest for the cases where the population canonical eigenvalues were equal or had little spread, i.e. matched the priors (heritability sets M, N, R and S and correlations  $I$  and  $IX$ ). However, overall improvements in estimates of  $\Sigma_G$  due to  $\mathcal{P}_\lambda$  were lowest among the penalties examined while the corresponding PRIAL in  $\Sigma_E$  was higher, especially for smallest sample size. Doubling the penalty ( $s = \frac{1}{2}$ ) (equivalent to shrinking both  $\lambda_i$  and  $1 - \lambda_i$ ) increased PRIAL for both  $\Sigma_G$  and  $\Sigma_E$ . Nevertheless, applying penalties after transformation to logarithmic scale resulted in higher PRIAL for  $\Sigma_G$  throughout. As observed previously (Meyer & Kirkpatrick, 2010; Meyer, 2011), penalizing both  $\log(\lambda_i)$  and  $\log(1 - \lambda_i)$  ( $\mathcal{P}_{\lambda\lambda}^+$ ) increased the PRIAL for  $\Sigma_E$  markedly without being detrimental to estimates of  $\Sigma_G$ . For the smaller samples this was best applied to the average of the two components (scale factor  $s = 2$ ) – as the high values for the mean change in (unpenalized) likelihood,  $\Delta \mathcal{L}$ , in Table 2 indicate, using the unscaled sum resulted in estimates of  $\nu$  at the lower limit, and thus too stringent a penalty, in several cases. Allowing for different values of  $\nu$  for the penalties on  $\log(\lambda_i)$  and  $\log(1 - \lambda_i)$  ( $\mathcal{P}_{\lambda\lambda*}^+$ ) did not prove advantageous whilst requiring

substantial additional computational effort for the two-dimensional optimisation.

Similarly, assuming a Beta distribution for the canonical eigenvalues yielded a penalty ( $\mathcal{P}_\beta$ ) targeting both  $\log(\lambda_i)$  and  $\log(1 - \lambda_i)$  simultaneously. As illustrated in Figure 1,  $\mathcal{P}_\beta$  resulted in a similar pattern of PRIAL as  $\mathcal{P}_{l\lambda}^+$ , except for the cases with little spread in the population canonical eigenvalues, with a lower level of PRIAL and  $\Delta\mathcal{L}$  overall. In part, the lower average could be attributed to estimates of  $\nu$  being consistently close to the lower limit allowed for a number of cases which led to very little regularisation. This behaviour indicated a shape of the profile likelihood unsuitable to direct estimation and led to the choice of a minimum value of  $\mu = 1$  rather than 0.01, which lifted improvements in estimates achieved to useful levels for all cases.

Initial attempts at using a penalty derived from an IW prior were less successful. As shown in Table 4 for a penalty on  $\Sigma_G$  only,  $\mathcal{P}_\Sigma$  as given in (Eq. 9), direct estimation of  $\nu$  resulted in rather large reductions in the unpenalized likelihood, especially for the smallest sample size. This was again due to a profile likelihood which did not allow adequate estimation of  $\nu$ , producing values close to the upper limit set for various sets of population values and thus dramatic over-shrinkage. As illustrated in Figure 1, low average PRIAL values hide substantial damage to estimates for a number of cases, mostly those for which penalties on the canonical eigenvalues worked best. Adding a penalty on  $\Sigma_E$  in addition did not alleviate the problem, regardless whether a joint ( $\mathcal{P}_\Sigma^+$ ) or different ( $\mathcal{P}_{\Sigma^*}^+$ ) degrees of freedom were estimated for the two components. In fact, PRIAL values for  $\Sigma_E$  resulting from  $\mathcal{P}_\Sigma^+$  and  $\mathcal{P}_{\Sigma^*}^+$  were, except for  $s = 1000$ , lower than for  $\mathcal{P}_\Sigma$ , suggesting that shrinkage towards the sum of estimates for  $\Sigma_G$  and  $\Sigma_E$  provided better regularization of the latter. When allowing for separate regulating factors, the mean of  $\nu$  for  $\Sigma_G$  was close to that for  $\mathcal{P}_\Sigma$  (see Table 2), while the mean value for  $\Sigma_E$  (not shown) ranged from 8.16 to 8.17, i.e. was close to the lowest possible value.

However, dividing the penalty by an arbitrary factor  $s$  to reduce its numerical impact dramatically improved the performance of this type of penalty. Though average PRIAL values did not increase much over a simple halving ( $s = 2$ ), higher scaling factors appeared advantageous as they reduced the variation across the different cases and ensured a positive PRIAL for  $\Sigma_E$  for all cases. This is illustrated in Figure 1 for  $s = 5$  (using  $\mathcal{P}_\Sigma^s$  to denote penalty  $\mathcal{P}_\Sigma$  for scaling factor  $s$ ). Moreover, they eliminated cases where over-penalization had distorted the partitioning of  $\Sigma_P$  into  $\Sigma_G$  and  $\Sigma_E$  to the extent that the new estimate of  $\Sigma_P$  had a higher loss than the estimate from a standard, unpenalized analysis. The latter phenomenon was clearly contrary to the aim of borrowing strength from the estimate of the phenotypic covariance matrix. For  $q = 9$ , a factor of  $s = 5$ , equivalent to  $(q + 1)/2$ , seemed a reasonable value.

#### 4.1.1 Comparison with previous strategies

Table 3 contrasts results achieved with DE for  $s = 100$  with those for  $q = 5$  and the subset of 30 cases for  $q = 9$  identical to those used in Meyer (2011). Overall, PRIAL values for DE for

$q = 5$  appeared comparable to those obtained previously using three-fold cross-validation to estimate a tuning factor regulating the strength of penalization. However, both tended to be somewhat lower than those achieved limiting the change in  $\log \mathcal{L}(\theta)$  to approximately  $-1.92$ . As values for  $\Delta \mathcal{L}$  for  $q = 9$  show, DE resulted in larger average changes for all but two penalties, i.e. on average DE in somewhat stronger penalization for the larger number of traits.

## 4.2 Fixed strength of penalization

Also shown in Table 2 are corresponding results for selected, fixed values of the regulating parameter. PRIAL values achieved in these analyses were generally less than those from DE but often represented a rather substantial proportion of the former, indicating that – with judicious choice of the ‘default’ parameter – this may be a computationally less demanding alternative worth considering.

Assuming a Normal distribution for the canonical eigenvalues on the logarithmic scale, under-penalisation by choosing variances two- or even four-fold that of the average obtained with direct estimation still yielded considerable average PRIAL values, especially for the smallest sample size. This suggests a welcome degree of robustness against sub-optimal choices for the strength of penalization and, moreover, a non-linear mode of action with initial, small modifications due to mild penalties achieving more ‘per unit’ than stronger regulating factors.

Figure 2 displays the distribution of PRIAL values across the 60 sets of population values considered for penalty  $\mathcal{P}_\beta$ , contrasting results from direct estimation of  $\nu$  with fixed values ranging from 1 to 20. While the PRIAL for  $\Sigma_G$  increased up to about  $\nu = 12$ , this was accompanied by increasing spread and rising proportion of cases for which values were negative for  $\Sigma_G$ ,  $\Sigma_E$  or  $\Sigma_P$ . Results shown suggest that a default value of  $\nu = 3$  or 4 may represent a good compromise between improving estimates of  $\Sigma_G$  and safe-guarding against ill-effects on the other components.

For penalty  $\mathcal{P}_\Sigma$ , a value for the degrees of freedom of the IW distribution between the minimum and  $q$ , appeared best (Table 4), regardless of sample size. This was markedly lower than the mean value from direct estimation and thus accompanied by much smaller changes in  $\log \mathcal{L}(\theta)$ , highlighting again that DE of  $\nu$  failed in a number of cases when it yielded estimates at or close to the upper limit imposed. Thus for fixed  $\nu$ , PRIAL in  $\Sigma_G$  changed relatively little over the range of scale factors considered,  $s = 1$  to  $s = 2q$ . However, PRIAL in  $\Sigma_E$  and  $\Delta \mathcal{L}$  decreased consistently with increasing  $s$ .

### 4.3 Less informative penalties

The other penalties considered were derived from priors devised for a single covariance, and did not include any feedback mechanism on the total variation available for partitioning into individual components as inherent in the penalties considered so far by targeting canonical eigenvalues or shrinking towards the phenotypic covariance matrix. Hence, application to  $\Sigma_G$  alone typically reduced losses for this matrix but, at the same time, resulted in increased losses in estimates of  $\Sigma_E$  and  $\Sigma_P$ . Hence only results for penalising both  $\Sigma_G$  and  $\Sigma_E$  simultaneously – using the same regulating factor for both if applicable – are reported in Table 2.

Like  $\mathcal{P}_\Sigma$ , the penalty proposed by Huang & Wand (2013),  $\mathcal{P}_{HW}$ , is based on the assumption of an IW distribution and thus involves a multiple of  $(\nu + q + 1)/2$  times the log determinant of the matrix estimate to be regularized. Not surprisingly then, scaling as for  $\mathcal{P}_\Sigma$  was required to achieve reasonable PRIAL values. Improvements using the suggested default of  $\nu = q + 1$  were relatively modest, with some increase for higher values. Similarly, penalty  $\mathcal{P}_J$  comprised a multiple of the log determinant proportional to the number of traits which, when applied without scaling, increased losses for estimates of  $\Sigma_G$  for a number of cases, together with a negative mean PRIAL for  $\Sigma_P$  and substantial changes in  $\log \mathcal{L}(\theta)$  for the smaller samples. Again, reducing the impact of the penalty by dividing by  $s > 1$  alleviated these problems.

On the other hand, the simple prior assuming a Wishart distribution yielded a penalty,  $\mathcal{P}_{Ch}$ , free of scaling problems with good average PRIAL values. As for  $\mathcal{P}_{HW}$ , using a slightly higher value than  $q + 2$ , suggested by Chung *et al.* (2013), for the degrees of freedom increased strength of penalization and improved estimates further. As noted above, for  $\nu = q + 2$ ,  $\mathcal{P}_{Ch}$  involves a term  $\frac{1}{2} \log |\Sigma|$  similar to  $\mathcal{P}_{R1}$ , and results for  $\mathcal{P}_{R1}$  were thus virtually identical to those for the former.

## 5 Discussion

Results clearly demonstrate that penalized maximum likelihood estimation for multivariate analyses in quantitative genetics can be highly advantageous, yielding estimates that are on average closer to the population values, especially for numerous traits and relatively small data sets. A range of penalties can be devised which ‘borrow strength’ from the estimate of the phenotypic covariance matrix, and thus redress problems arising from sampling variation in the partitioning of the total variation into its causal components. As reported previously (Meyer, 2011), these have somewhat different modes of action with none ‘best’ across the wide range of population values considered.

The crucial decision in regularized estimation is the choice of the strength of penalization. Attempts to determine an optimal value for a particular analysis can increase computational requirements substantially. Moreover, they have been found to be ‘noisy’ and afflicted by

problems of locating the maximum of the penalized likelihood with sufficient accuracy, in particular for smaller data sets or higher dimensional analyses. For direct estimation, these resulted in profile likelihood functions for  $\nu$  that were not quite smooth which, in turn, limited the scope for estimating  $\nu$  accurately. This held for both the quadratic approximation used here and a more laborious sequential search – for comparison the latter was carried out for selected analyses, but generally did not yield better PRIAL values. Similiar problems were encountered previously when trying to estimate a tuning factor.

As shown, direct estimation tended to work best for the REML analogue of ‘bending’, i.e. shrinkage of the (canonical) eigenvalues towards their mean. Considering a sample covariance matrix, Ledoit & Wolf (2004) presented an estimator for the amount to shrink this matrix towards a multiple of an identity matrix – equivalent to shrinking sample eigenvalues towards their mean – which minimized the quadratic loss of the resulting, modified estimate. Extensions to other shrinkage targets were considered, for instance, by Schäfer & Strimmer (2005) and Fisher & Sun (2011). Existence of such estimators is encouraging for the suitability of DE to determine the variance of the assumed Normal distribution of eigenvalues for penalties  $\mathcal{P}_\lambda$ ,  $\mathcal{P}_{l\lambda}$  or  $\mathcal{P}_{l\lambda}^+$ .

When using a tuning factor to determine the strength of penalization, penalties derived assuming an IW distribution of the matrix tended to yield comparable PRIAL values to those targeting the canonical eigenvalues. With DE, however, the resulting degree of penalization tended to be too strong owing to very high estimates of the degrees of freedom, suggesting a profile likelihood for  $\nu$  without a pronounced peak. While scaling the penalty to reduce its impact proved successful, this involved an *ad hoc* choice, and we are not aware of theoretical grounds to justify the value used. Fortunately, PRIAL for the estimates of  $\Sigma_G$  appeared relatively stable over a range of scaling factors. Presumably, the tuning factor in earlier work (Meyer, 2011) automatically included any scaling necessary.

Difficulties in estimating the strength of shrinkage together with the computational effort involved have limited applications of penalized REML estimation so far, and raise the question as to whether it is worthwhile in routine applications. Alternatively, ‘default’ values inducing a mild penalty might yield a substantial proportion of the benefits achievable for negligible additional effort and, moreover, reduce the risk of damage due to over-penalization. These penalties would be most effective for smaller samples and have little impact for large data sets. This is akin to the ideas underlying the use of weakly informative priors in Bayesian estimation, discussed above. Our results suggest that suitable values might be  $\nu = 1.0$  for the variance of canonical eigenvalues on the logarithmic scale for penalty  $\mathcal{P}_{l\lambda}^+$ ,  $\nu = 3$  to 4 for the effective number of prior observations for penalty  $\mathcal{P}_\beta$ , and  $\nu = q + 0.1$  to  $\nu = q + 1$  for the degrees of freedom in the IW distribution. In addition, even a simple, parameter-free penalty comprised of the log determinant of the matrix estimate,  $-\frac{1}{2} \log |\Sigma|$ , proved remarkably effective, and may be well worth considering.

All penalties considered improved estimates of covariance matrices by ‘lifting’ the smallest



eigenvalues. Eigenvalues close to zero are a major cause of slow or non-convergence in multivariate REML analyses, and thus excessive computational requirements. While reduced rank analyses have the scope to improve convergence rates, they are not without problems, as they may pick up the wrong set of principal components (i.e. the reduced rank estimator is not consistent; Meyer & Kirkpatrick (2008)). Hence, while not examined here, penalized estimation which steers eigenvalues away from zero may offer additional benefits by improving convergence of iterative REML analyses.

## 6 Conclusions

Penalized maximum likelihood estimation of genetic covariance matrices can dramatically reduce sampling variation and thus ‘improve’ estimates. It should be applied on a routine basis for analyses considering more than a handful of traits. The penalty can be formulated so that the hyper parameters of the assumed prior distribution regulate the strength of penalization. While this provides the opportunity to estimate these directly from the data at hand, suitable ‘default’ values can be identified which allow a substantial proportion of improvements feasible to be obtained with little effort, which may facilitate increased uptake of the methodology proposed.

## Acknowledgements

This work was supported by Meat and Livestock Australia under grant B.BFG.0050. I am indebted to Gustavo los Campos for suggesting to directly estimate the parameters of the assumed prior distribution.

## References

- Barnard J., McCulloch R., Meng X. (2000) Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Stat. Sin.*, **10**, 1281–1312.
- Bickel P. J., Levina E. (2008) Regularized estimation of large covariance matrices. *Ann. Stat.*, **36**, 199–227. doi: [10.1214/009053607000000758](https://doi.org/10.1214/009053607000000758).
- Bien J., Tibshirani R. J. (2011) Sparse estimation of a covariance matrix. *Biometrika*, **98**, 807–820. doi: [10.1093/biomet/asr054](https://doi.org/10.1093/biomet/asr054).
- Bouriga M., Féron O. (2013) Estimation of covariance matrices based on hierarchical inverse-Wishart priors. *J. Statist. Plann. Inference*, **143**, 795–808. doi: [10.1016/j.jspi.2012.09.006](https://doi.org/10.1016/j.jspi.2012.09.006).



- Brown P. J. (2006) Inverted Wishart Distribution, Generalized. In: *Encyclopedia of Environ-*  
*metrics*, John Wiley and Sons, Ltd. pp. 1079–1082. doi: 10.1002/9780470057339.vag008.
- Chung Y., Gelman A., Rabe-Hesketh S., Liu J., Dorie V. (2013) Weakly informative prior for  
 point estimation of covariance matrices in hierarchical models. URL [www.stat.columbia.](http://www.stat.columbia.edu/~gelman/research/unpublished/ms.pdf)  
[edu/~gelman/research/unpublished/ms.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/ms.pdf).
- Daniels M. J., Kass R. E. (1999) Nonconjugate Bayesian estimation of covariance matrices  
 and its use in hierarchical models. *J. Amer. Stat. Ass.*, **94**, 1254–1263.
- Daniels M. J., Kass R. E. (2001) Shrinkage estimators for covariance matrices. *Biometrics*, **57**,  
 1173–1184.
- Evans M., Hastings N., Peacock B. (2000) Beta distribution. In: *Statistical distributions*, Series  
 in Probability and Statistics, Wiley, New York, chap. 5. 3rd edn., pp. 34–42.
- Fisher T. J., Sun X. (2011) Improved Stein-type shrinkage estimators for the high-dimensional  
 multivariate normal covariance matrix. *Comp. Stat. Dat. Anal.*, **55**, 1909–1918. doi:  
 10.1016/j.csda.2010.12.006.
- Gaskins J. T., Daniels M. J., Marcus B. H. (2013) Sparsity inducing prior distributions for  
 correlation matrices of longitudinal data. *J. Comp. Graph. Stat.*, **0**, 000–000, published  
 online October 30, 2013. doi: 10.1080/10618600.2013.852553.
- Gelman A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian*  
*analysis*, **1**, 515–533. doi: 10.1214/06-BA117A.
- Gelman A., Hill J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*.  
 Cambridge University Press.
- Harville D. A. (1977) Maximum likelihood approaches to variance component estimation  
 and related problems. *J. Amer. Stat. Ass.*, **72**, 320–338.
- Hayes J. F., Hill W. G. (1981) Modifications of estimates of parameters in the construction of  
 genetic selection indices ('bending'). *Biometrics*, **37**, 483–493.
- Hsu C. W., Sinay M. S., Hsu J. S. J. (2010) Bayesian estimation of a covariance matrix with  
 flexible prior specification. *Ann. Inst. Statist. Math.*, 1–24. doi: 10.1007/s10463-010-0314-5.
- Huang A., Wand M. P. (2013) Simple marginally noninformative prior distributions for  
 covariance matrices. *Bayesian Analysis*, **8**, 439–452. doi: 10.1214/13-BA815.
- Huang J. Z., Liu N., Pourahmadi M., Liu L. (2006) Covariance matrix selection and estimation  
 via penalised normal likelihood. *Biometrika*, **93**, 85–98. doi: 10.1093/biomet/93.1.85.
- James W., Stein C. (1961) Estimation with quadratic loss. In: Neiman J. (ed.), *Proceedings of the*  
*Fourth Berkeley Symposium on Mathematical Statistics and Probability*. University of California  
 Press, pp. 361–379.

- Lawley D. N. (1956) Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, **43**, 128–136. doi: [10.1093/biomet/43.1-2.128](https://doi.org/10.1093/biomet/43.1-2.128).
- Ledoit O., Wolf M. (2004) A well-conditioned estimator for large-dimensional covariance matrices. *J. Multiv. Anal.*, **88**, 365–411. doi: [10.1016/S0047-259X\(03\)00096-4](https://doi.org/10.1016/S0047-259X(03)00096-4).
- Ledoit O., Wolf M. (2012) Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, **40**, 1024–1060. doi: [10.1214/12-AOS989](https://doi.org/10.1214/12-AOS989).
- Leonard T. A., Hsu J. S. J. (1992) Bayesian inference for a covariance matrix. *Ann. Stat.*, **20**, 1669–1696.
- Lin S. P., Perlman M. D. (1985) A Monte Carlo comparison of four estimators of a covariance matrix. In: Krishnaish P. R. (ed.), *Multivariate Analysis*, vol. 6. North-Holland, Amsterdam, vol. 6, pp. 411–428.
- Meyer K. (2011) Performance of penalized maximum likelihood in estimation of genetic covariances matrices. *Genet. Sel. Evol.*, **43**, 39. doi: [10.1186/1297-9686-43-39](https://doi.org/10.1186/1297-9686-43-39).
- Meyer K. (2013) Penalized estimation of covariance matrices with flexible amounts of shrinkage. *Proc. Ass. Advan. Anim. Breed. Genet.*, **20**, 428–431.
- Meyer K., Kirkpatrick M. (2008) Perils of parsimony: Properties of reduced rank estimates of genetic covariances. *Genetics*, **108**, 1153–1166. doi: [10.1534/genetics.108.090159](https://doi.org/10.1534/genetics.108.090159).
- Meyer K., Kirkpatrick M. (2010) Better estimates of genetic covariance matrices by ‘bending’ using penalized maximum likelihood. *Genetics*, **185**, 1097–1110. doi: [10.1534/genetics.109.113381](https://doi.org/10.1534/genetics.109.113381).
- Meyer K., Kirkpatrick M., Gianola D. (2011) Penalized maximum likelihood estimates of genetic covariance matrices with shrinkage towards phenotypic dispersion. *Proc. Ass. Advan. Anim. Breed. Genet.*, **19**, 87–90.
- Powell M. (2006) The NEWUOA software for unconstrained optimization without derivatives. In: Pillo G., Roma M. (eds.), *Large-Scale Nonlinear Optimization*, vol. 83 of *Nonconvex Optimization and Its Applications*. ISBN 978-0-387-30063-4, pp. 255–297. doi: [10.1007/0-387-30065-1\\_16](https://doi.org/10.1007/0-387-30065-1_16).
- Rothman A. J., Levina E., Zhu J. (2010) A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika*, **97**, 539–550. doi: [10.1093/biomet/asq022](https://doi.org/10.1093/biomet/asq022).
- Schäfer J., Strimmer K. (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.*, **4**, 32. doi: [10.2202/1544-6115.1175](https://doi.org/10.2202/1544-6115.1175).
- Stein C. (1975) Estimation of a covariance matrix. In: *Reitz lecture of the 39th Annual Meeting of the Institute of Mathematical Statistics*. Atlanta.

- Sun D., Berger J. (2007) Objective Bayesian analysis for the multivariate normal model. In: Bernardo J., Bayarri M., Berger J., Dawid A., Heckerman D., Smith A., West M. (eds.), *Bayesian Statistics*, vol. 8, Oxford University Press. pp. 525–547.
- Warton D. I. (2008) Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *J. Amer. Stat. Ass.*, **103**, 340–349. doi: [10.1198/016214508000000021](https://doi.org/10.1198/016214508000000021).
- Witten D. M., Tibshirani R. (2009) Covariance-regularized regression and classification for high dimensional problems. *J. Roy. Stat. Soc. B*, **71**, 615–636. doi: [10.1111/j.1467-9868.2009.00699.x](https://doi.org/10.1111/j.1467-9868.2009.00699.x).
- Won J.-H., Lim J., Kim S.-J., Rajaratnam B. (2013) Condition-number-regularized covariance estimation. *J. Roy. Stat. Soc. B*, **75**, 427–450. doi: [10.1111/j.1467-9868.2012.01049.x](https://doi.org/10.1111/j.1467-9868.2012.01049.x).
- Yang R., Berger J. O. (1994) Estimation of a covariance matrix using the reference prior. *Ann. Stat.*, 1195–1211.
- Ye R. D., Wang S. G. (2009) Improved estimation of the covariance matrix under Stein’s loss. *Stat. Prob. Lett.*, **79**, 715 – 721. doi: [10.1016/j.spl.2008.10.024](https://doi.org/10.1016/j.spl.2008.10.024).

## A Details for population parameters

Heritabilities, summarized in Table 5, were chosen to represent a range of cases, from equal values for all traits to sets of values which not only spanned almost the entire interval from zero to unity but also were very unevenly distributed. Correlation matrices selected included compound symmetric and auto-regressive structures, with genetic correlations up to 0.8. Table 6 gives details of the values used together with the assumed phenotypic variances, as a function of trait numbers. Combining heritabilities and correlation scenarios yielded sets of covariance matrices with coefficients of variation among the corresponding canonical eigenvalues ranging from 0 to 175% for  $q = 5$  and 0 to 150% for  $q = 9$ . Population values for  $q = 5$  were identical to those used previously, and M-I to V-I, M- $\mathcal{V}$ I to V- $\mathcal{V}$ I and M- $\mathcal{V}$ II to V- $\mathcal{V}$ II matched the 30 cases for  $q = 9$  considered in Meyer (2011).

Table 1: Summary of penalties fitted with scaling factors and fixed values for regulation parameters used.

Prior <sup>a</sup>	Penalty	Target	RP <sup>b</sup>	$f(q)$ <sup>c</sup>	DE <sup>d</sup>	Fixed RP <sup>e</sup>	Scale factors <sup>f</sup>	Eq. <sup>g</sup>
Normal	$\mathcal{P}_\lambda$	$\lambda$	1	0	✓	0.04 to 0.08	0.5	(4)
log-Normal	$\mathcal{P}_{l\lambda}$	$\log(\lambda)$	1	0	✓	–	–	(4)
	$\mathcal{P}_{l\lambda}^+$	$\log(\lambda), \log(1-\lambda)$	1	0	✓	0.5 to 2	2	(4)
	$\mathcal{P}_{l\lambda}^{+*}$	$\log(\lambda), \log(1-\lambda)$	2	0	✓	–	2	(4)
Beta	$\mathcal{P}_\beta$	$\log(\lambda), \log(1-\lambda)$	1	0	✓	1 to 8, 12, 16, 20	–	(5)
Inv. Wishart	$\mathcal{P}_\Sigma$	$\Sigma_G$	1	$q-1$	✓	0.1, 1 to 9	$2, q^{\frac{1}{2}}, \frac{q+1}{2}, q, 2q$	(9)
	$\mathcal{P}_\Sigma^+$	$\Sigma_G, \Sigma_E$	1	$q-1$	✓	as $\mathcal{P}_\Sigma$	as $\mathcal{P}_\Sigma$	(9)
	$\mathcal{P}_{\Sigma^+}^+$	$\Sigma_G, \Sigma_E$	2	$q-1$	✓	as $\mathcal{P}_\Sigma$	as $\mathcal{P}_\Sigma$	(9)
	$\mathcal{P}_{HW}$	$\Sigma_G, \Sigma_E$	1	$q-1$	×	2 to 6, 8	as $\mathcal{P}_\Sigma$	(13)
Wishart	$\mathcal{P}_{Ch}$	$\Sigma_G, \Sigma_E$	1	$q$	×	as $\mathcal{P}_{HW}$	–	(15)
Jeffreys	$\mathcal{P}_J$	$\Sigma_G, \Sigma_E$	0	–	×	–	as $\mathcal{P}_\Sigma$	(17)
Reference	$\mathcal{P}_{R1}$	$\Sigma_G, \Sigma_E$	0	–	×	–	–	(18)

<sup>a</sup>Prior or distribution used to derive penalty<sup>b</sup>Number of regulating parameters<sup>c</sup>Function of number of traits in formulation of  $\nu$ <sup>d</sup>Direct estimation of regulating parameters: ✓ yes, × no<sup>e</sup>Constants for  $\mu + \exp(\kappa)$  used to set fixed regulating parameter<sup>f</sup>Scale factors used<sup>g</sup>Pertaining equation number in section 2

Table 2: Mean percentage reduction in average loss for estimates of genetic ( $\Sigma_G$ ), residual ( $\Sigma_E$ ) and phenotypic ( $\Sigma_P$ ) covariance matrices together with mean regulating parameter ( $\nu$ ) and mean change in unpenalised log likelihood from the maximum ( $\Delta\mathcal{L}$ ), summarizing selected cases.

Penalty <sup>a</sup>		100 sires					400 sires					1000 sires				
	$s^b$	$\nu^c$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\nu$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\nu$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$
<i>Estimated regulating parameters</i>																
$\mathcal{P}_\lambda$	1	4.94	14	47	1	-0.99	4.49	12	21	1	-0.34	4.37	7	10	0	-0.14
	$\frac{1}{2}$	3.23	37	54	1	-5.16	4.03	23	28	1	-1.37	4.18	14	14	0	-0.54
$\mathcal{P}_{l\lambda}$	1	0.75	64	26	1	-3.65	0.81	57	14	0	-1.69	0.82	44	9	0	-0.92
$\mathcal{P}_{l\lambda}^+$	1	0.17	60	53	2	-16.51	0.37	61	37	1	-5.75	0.43	49	23	1	-2.59
	2	0.51	65	54	1	-3.45	0.52	56	30	1	-1.54	0.51	43	18	0	-0.82
$\mathcal{P}_{l\lambda}^+$	1	0.34	63	50	1	-16.13	0.71	61	33	1	-4.24	0.78	47	19	0	-1.91
	2	1.09	62	56	1	-2.70	1.02	52	31	1	-1.13	0.94	38	18	0	-0.56
$\mathcal{P}_\beta$	1	4.74	58	52	1	-2.13	4.93	45	26	1	-0.77	4.79	31	13	0	-0.33
$\mathcal{P}_\Sigma$	5	13.60	57	25	0	-2.63	12.18	46	9	0	-1.25	11.87	34	4	0	-0.77
$\mathcal{P}_\Sigma^+$	5	9.27	55	1	-1	-2.02	9.09	46	5	0	-1.14	9.06	34	4	0	-0.77
$\mathcal{P}_{\Sigma^+}^+$	5	13.02	55	10	-1	-2.19	12.12	45	8	0	-1.13	11.88	33	5	0	-0.71
<i>Fixed regulating parameters</i>																
$\mathcal{P}_\lambda$	$\frac{1}{2}$	4.00	24	52	1	-2.97	4.00	17	23	0	-1.12	4.00	10	12	0	-0.51
		6.00	17	52	1	-1.70	6.00	12	25	1	-0.55	6.00	7	12	0	-0.24
		8.00	14	50	1	-1.11	8.00	9	23	1	-0.33	8.00	6	11	0	-0.14
$\mathcal{P}_{l\lambda}^+$	2	0.50	62	49	1	-2.66	0.50	50	25	0	-1.35	0.50	38	14	0	-0.77
		1.00	58	44	1	-1.53	1.00	46	22	0	-0.72	1.00	34	12	0	-0.39
		2.00	52	38	1	-0.88	2.00	41	18	0	-0.39	2.00	29	10	0	-0.19
$\mathcal{P}_\beta$	1	4.00	61	53	1	-2.53	4.00	47	26	1	-1.04	4.00	34	14	0	-0.52
$\mathcal{P}_\Sigma$	5	9.00	58	22	0	-2.18	9.00	46	9	0	-1.22	9.00	35	4	0	-0.80
$\mathcal{P}_\Sigma^+$	5	9.00	55	3	-1	-2.03	9.00	46	7	0	-1.16	9.00	34	4	0	-0.77
<i>Penalties based on weakly informative priors</i>																
$\mathcal{P}_{HW}$	5	10.00	6	28	1	-3.84	10.00	34	5	0	-0.79	10.00	23	-2	0	-0.33
$\mathcal{P}_{HW}$	5	12.00	17	32	2	-4.02	12.00	38	8	0	-0.85	12.00	26	1	0	-0.37
$\mathcal{P}_{HW}$	9	10.00	30	31	1	-1.24	10.00	35	9	0	-0.38	10.00	23	2	0	-0.18
$\mathcal{P}_{HW}$	9	12.00	40	34	1	-1.35	12.00	38	11	0	-0.44	12.00	25	4	0	-0.20
$\mathcal{P}_{Ch}$	1	11.00	51	36	1	-1.03	11.00	40	18	0	-0.42	11.00	29	9	0	-0.20
$\mathcal{P}_{Ch}$	1	13.00	56	43	1	-3.35	13.00	48	23	0	-1.53	13.00	36	13	0	-0.81
$\mathcal{P}_{Ch}$	1	15.00	53	46	0	-5.83	15.00	48	24	0	-2.84	15.00	37	14	0	-1.59
<i>No regulating parameter</i>																
$\mathcal{P}_J$	1	–	38	47	-4	-12.33	–	40	24	-1	-6.63	–	32	14	0	-4.00
$\mathcal{P}_J$	5	–	56	41	1	-2.17	–	46	21	0	-0.94	–	34	11	0	-0.48
$\mathcal{P}_{R1}$	1	–	51	36	1	-1.03	–	40	18	0	-0.42	–	28	8	0	-0.20

<sup>a</sup>see text for acronyms

<sup>b</sup>Scale factor

<sup>c</sup>×100 for  $\mathcal{P}_\lambda$

Table 3: Mean percentage reduction in average loss for estimates of genetic ( $\Sigma_G$ ), residual ( $\Sigma_E$ ) and phenotypic ( $\Sigma_P$ ) covariance matrices for different strategies to determine the extent of penalisation.

Data for  $s = 100$  sires

Cov. <sup>a</sup>	Strategy <sup>b</sup>	Penalty <sup>c</sup>							
		$\mathcal{P}_\lambda$	$\mathcal{P}_{l\lambda}$	$\mathcal{P}_{l\lambda}^+$	$\mathcal{P}_{l\lambda^*}^+$	$\mathcal{P}_\beta$	$\mathcal{P}_\Sigma$	$\mathcal{P}_\Sigma^+$	$\mathcal{P}_{\Sigma^*}^+$
5 traits									
$\Sigma_G$	DE	17	62	63	63	55	62	57	60
	CV3	23	56	61	–	58	55	53	–
	L5%	41	68	70	–	69	64	67	–
$\Sigma_E$	DE	40	24	52	51	35	26	29	36
	CV3	14	27	44	–	36	11	43	–
	L5%	44	35	56	–	54	7	51	–
$\Sigma_P$	DE	0.7	0.5	0.9	0.7	-0.3	0.1	-0.4	-0.3
	CV3	-0.4	0.4	0.5	–	0.3	0.2	0.1	–
	L5%	-0.7	0.7	0.8	–	0.5	0.3	1.0	–
9 traits - subset									
$\Sigma_G$	DE	11	58	58	57	52	56	52	53
	L5%	24	68	68	–	67	68	68	–
$\Sigma_E$	DE	47	23	58	58	55	25	10	17
	L5%	63	16	59	–	63	10	47	–
$\Sigma_P$	DE	1.0	0.8	1.4	1.4	1.1	0.3	-1.1	-0.9
	L5%	1.2	0.5	1.1	–	1.2	0.6	0.7	–
$\Delta\mathcal{L}$	DE	-0.84	-3.11	-2.94	-2.47	-1.55	-2.59	-2.12	-2.21

<sup>a</sup>Covariance matrix

<sup>b</sup>DE: direct estimation, CV3: three-fold cross-validation and L5%: limiting the change in likelihood to  $\approx -1.92$ ; results for CV3 and L5% from Meyer (2011; Table 1)

<sup>c</sup>see text; using  $s = 2$  for  $\mathcal{P}_{l\lambda}^+$  and  $\mathcal{P}_{l\lambda^*}^+$ , and  $s = (q + 1)/2$  for  $\mathcal{P}_\Sigma$ ,  $\mathcal{P}_\Sigma^+$  and  $\mathcal{P}_{\Sigma^*}^+$

Table 4: Mean percentage reduction in average loss (PRIAL) together with mean associated statistics for a penalty on  $\Sigma_G$  based on the Inverse Wishart distribution ( $\mathcal{P}_\Sigma$ ), employing different scaling factors ( $s$ ) and values for the degrees of freedom ( $\nu$ ).

$\kappa$	100 sires					400 sires					1000 sires				
$s$	1	2	5	9	18	1	2	5	9	18	1	2	5	9	18
<i>PRIAL for genetic covariance matrix</i>															
DE <sup>a</sup>	14	50	57	55	52	26	44	46	44	42	24	33	34	32	31
8.1	60	62	59	56	53	42	47	47	45	43	31	35	35	34	32
9.0	58	61	58	55	52	40	46	46	45	43	29	34	35	33	32
11.0	53	59	58	55	52	34	44	45	44	42	26	33	34	33	31
13.0	47	56	56	54	52	27	41	44	44	42	20	30	33	32	30
17.0	33	49	54	53	50	8	34	42	42	40	6	24	31	30	29
<i>PRIAL for residual covariance matrix</i>															
DE	-4	33	25	16	9	5	14	9	6	3	4	6	4	3	2
8.1	44	39	22	14	8	18	15	8	5	3	8	7	4	3	1
9.0	42	39	22	14	8	17	15	9	5	3	7	7	4	2	2
11.0	39	40	24	15	8	15	15	9	6	3	7	7	4	3	2
13.0	34	40	25	16	9	12	15	9	6	3	4	7	4	3	2
17.0	22	39	27	17	10	3	14	9	6	4	0	6	4	3	2
<i>Mean change in log likelihood</i>															
DE	-47.1	-14.0	-2.6	-1.5	-1.0	-13.0	-3.6	-1.3	-0.8	-0.6	-5.3	-1.9	-0.8	-0.5	-0.3
8.1	-13.9	-5.4	-2.1	-1.4	-1.0	-6.2	-2.6	-1.2	-0.9	-0.6	-3.5	-1.6	-0.8	-0.5	-0.4
9.0	-15.4	-5.8	-2.2	-1.5	-1.0	-6.8	-2.8	-1.2	-0.9	-0.6	-3.8	-1.7	-0.8	-0.5	-0.4
11.0	-19.1	-6.9	-2.3	-1.5	-1.0	-8.4	-3.1	-1.2	-0.8	-0.6	-4.5	-1.8	-0.8	-0.5	-0.3
13.0	-23.3	-8.3	-2.5	-1.5	-1.0	-10.4	-3.6	-1.3	-0.8	-0.6	-5.4	-2.0	-0.8	-0.5	-0.3
17.0	-33.2	-11.8	-3.1	-1.6	-1.0	-15.5	-5.0	-1.5	-0.9	-0.5	-7.9	-2.6	-0.9	-0.5	-0.3
<i>Mean value for direct estimates of <math>\nu</math></i>															
DE	23.6	18.6	13.6	13.1	13.0	14.7	12.7	12.2	12.2	12.2	12.4	12.0	11.9	11.9	11.9

<sup>a</sup>Directly estimated value



Table 5: Population values ( $\times 100$ ) for sets of heritabilities

No. <sup>a</sup>	A	B	C	D	E	F	G	H	I	J	K	L
1	40	50	60	70	80	90	20	30	60	50	90	70
2	40	45	50	55	30	50	20	25	10	20	30	70
3	40	40	40	40	30	30	20	20	10	15	10	40
4	40	35	30	25	30	20	20	15	10	10	10	10
5	40	30	20	10	30	10	20	10	10	5	10	10
	M	N	O	P	Q	R	S	T	U	V	W	X
1	40	60	90	75	70	20	35	60	50	80	30	35
2	40	55	60	70	70	20	30	50	50	40	30	30
3	40	50	50	60	70	20	25	10	20	10	25	30
4	40	45	50	50	40	20	20	10	15	10	25	20
5	40	40	30	40	40	20	20	10	15	10	20	20
6	40	35	30	30	40	20	20	10	10	10	15	15
7	40	30	20	20	10	20	15	10	10	10	15	15
8	40	25	20	10	10	20	10	10	5	5	10	15
9	40	20	10	5	10	20	5	10	5	5	10	10

<sup>a</sup>Trait number

Table 6: Population values for correlations and phenotypic variances

Case	$q^a$	Correlations <sup>b</sup>		Variances <sup>c</sup>
		Genetic	Residual	
<i>I</i>	5, 9	0	0	1
<i>II</i>	5	0.8	0	$1.5^{i-1}$
<i>III</i>	5	$0.6^{ i-j }$	$-0.4^{ i-j } + 0.5$	as $\mathcal{V}$
<i>IV</i>	5	$-0.8^{ i-j } + 0.02i$	$-0.4^{ i-j } + 0.5$	as $\mathcal{V}$
$\mathcal{V}$	5	$-1^i 0.05j + 0.5$	$-1^j 0.1i + 0.2$	$1 +  3 - i $
$\mathcal{V}I$	9	$0.7^{ i-j }$	$-1^j 0.05i + 0.2$	as <i>IX</i>
$\mathcal{V}II$	9	$-0.8^{ i-j } + 0.02i$	$-0.2^{ i-j } + 0.5$	as <i>IX</i>
$\mathcal{V}III$	9	0.5	0.3	as <i>IX</i>
<i>IX</i>	9	0.7 if $i, j \in [3, 7]$ 0.3 otherwise	0.7 if $i, j \in [3, 7]$ 0.3 otherwise	$1 +  5 - i $ if $i \in [3, 7]$ mod( $i, 7$ ) otherwise

<sup>a</sup>Number of traits for which applied<sup>b</sup>between traits  $i$  and  $j$  ( $i \neq j$ )<sup>c</sup>Phenotypic variance for trait  $i$

Figure 1: Percentage reduction in average loss for estimates of the genetic and residual covariance matrix for individual sets of population values and selected penalties (see text for acronyms)

Data for  $q = 9$  traits and  $s = 400$  sires

	P <sub>λ</sub>									P <sub>Iλ</sub>									P <sub>I<sub>λ</sub></sub> <sup>+</sup>									P <sub>β</sub>									P <sub>Σ</sub>									P <sub>Σ</sub> <sup>5</sup>																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
	30	6	4	28	28	31	5	4	27	29	59	64	62	69	60	62	68	58	64	62	59	70	63	63	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62	59	70	63	69	68	58	64	62

Correlations

Figure 2: Distribution of the percentage reduction in average loss due to a penalty assuming a Beta distribution of the canonical eigenvalues ( $\mathcal{P}_\beta$ ), for directly estimated (DE) and fixed values for the effective prior size

Data for  $q = 9$  traits and considering samples with  $s = 100$ ,  $s = 400$  and  $s = 1000$  sires.

