

# Simple Penalties on Maximum-Likelihood Estimates of Genetic Parameters to Reduce Sampling Variation

Karin Meyer<sup>1</sup>

Animal Genetics and Breeding Unit, University of New England, Armidale, New South Wales 2351, Australia

ORCID ID: 0000-0003-2663-9059 (K.M.)

**ABSTRACT** Multivariate estimates of genetic parameters are subject to substantial sampling variation, especially for smaller data sets and more than a few traits. A simple modification of standard, maximum-likelihood procedures for multivariate analyses to estimate genetic covariances is described, which can improve estimates by substantially reducing their sampling variances. This is achieved by maximizing the likelihood subject to a penalty. Borrowing from Bayesian principles, we propose a mild, default penalty—derived assuming a Beta distribution of scale-free functions of the covariance components to be estimated—rather than laboriously attempting to determine the stringency of penalization from the data. An extensive simulation study is presented, demonstrating that such penalties can yield very worthwhile reductions in loss, *i.e.*, the difference from population values, for a wide range of scenarios and without distorting estimates of phenotypic covariances. Moreover, mild default penalties tend not to increase loss in difficult cases and, on average, achieve reductions in loss of similar magnitude to computationally demanding schemes to optimize the degree of penalization. Pertinent details required for the adaptation of standard algorithms to locate the maximum of the likelihood function are outlined.

**KEYWORDS** genetic parameters; improved estimates; regularization; maximum likelihood; penalty

**E**STIMATION of genetic parameters, *i.e.*, partitioning of phenotypic variation into its causal components, is one of the fundamental tasks in quantitative genetics. For multiple characteristics of interest, this involves estimation of covariance matrices due to genetic, residual, and possibly other random effects. It is well known that such estimates can be subject to substantial sampling variation. This holds especially for analyses comprising more than a few traits, as the number of parameters to be estimated increases quadratically with the number of traits considered, unless the covariance matrices of interest have a special structure and can be modeled more parsimoniously. Indeed, a sobering but realistic view is that “Few datasets, whether from livestock, laboratory or natural populations, are of sufficient size to obtain useful estimates of many genetic parameters” (Hill 2010, p. 75). This not only emphasizes the importance of appropriate

data, but also implies that a judicious choice of methodology for estimation—which makes the most of limited and precious records available—is paramount.

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. In most cases, this involves reducing sampling variance at the expense of some bias—if the additional bias is small and the reduction in variance sufficiently large, the loss is reduced. In statistical parlance “regularization” refers to the use of some kind of additional information in an analysis. This is often used to solve ill-posed problems or to prevent overfitting through some form of penalty for model complexity; see Bickel and Li (2006) for a review. There has been longstanding interest, dating back to Stein (1975) and earlier (James and Stein 1961), in regularized estimation of covariance matrices to reduce their loss. Recently, as estimation of higher-dimensional matrices is becoming more ubiquitous, there has been a resurgence in interest (*e.g.*, Bickel and Levina 2008; Warton 2008; Witten and Tibshirani 2009; Ye and Wang 2009; Rothman *et al.* 2010; Fisher and Sun 2011; Ledoit and Wolf 2012; Deng and Tsui 2013; Won

Copyright © 2016 by the Genetics Society of America

doi: 10.1534/genetics.115.186114

Manuscript received December 14, 2015; accepted for publication June 9, 2016; published Early Online June 13, 2016.

Supplemental material is available online at [www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.186114/-/DC1](http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.186114/-/DC1).

<sup>1</sup>Address for correspondence: Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia. E-mail: [kmeyer@une.edu.au](mailto:kmeyer@une.edu.au)

*et al.* 2013). In particular, estimation encouraging sparsity is an active field of research for estimation of covariance matrices (e.g., Pourahmadi 2013) and in related areas, such as graphical models and structural equations.

### Improving Estimates of Genetic Parameters

As emphasized above, quantitative genetic analyses require at least two covariance matrices to be estimated, namely due to additive genetic and residual effects. The partitioning of the total variation into its components creates substantial sampling correlations between them and tends to exacerbate the effects of sampling variation inherent in estimation of covariance matrices. However, most studies on regularization of multivariate analyses considered a single covariance matrix only and the literature on regularized estimates of more than one covariance matrix is sparse. In a classic article, Hayes and 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 ( $\Sigma_P$ ) toward their mean, a procedure described as “bending” the estimate of  $\Sigma_G$  toward that of  $\Sigma_P$ . The underlying rationale was that  $\Sigma_P$ , the sum of all the causal components, is typically estimated much more accurately than any of its components, so that bending would “borrow strength” from the estimate of  $\Sigma_P$ , while shrinking estimated eigenvalues toward their mean would counteract their known, systematic overdispersion. The authors demonstrated by simulation that use of “bent” estimates in constructing selection indexes could increase the achieved response to selection markedly, as these were closer to the population values than unmodified estimates and thus provided more appropriate estimates of index weights. However, no clear guidelines to determine the optimal amount of shrinkage to use were available and bending was thus primarily used only to modify nonpositive definite estimates of covariance matrices and all but forgotten when methods that allowed estimates to be constrained to the parameter space became common procedures.

Modern analyses to estimate genetic parameters are generally carried out fitting a linear mixed model and using restricted maximum-likelihood (REML) or Bayesian methodology. The Bayesian framework directly offers the opportunity for regularization through the choice of appropriate priors. Yet, this is rarely exploited for this purpose and “flat” or minimally informative priors are often used instead (Thompson *et al.* 2005). In a maximum-likelihood context, estimates can be regularized by imposing a penalty on the likelihood function aimed at reducing their sampling variance. This provides a direct link to Bayesian estimation: For a given prior distribution of the parameters of interest or functions thereof, an appropriate penalty can be obtained as a multiple of minus the logarithmic value of the probability density function. For instance, shrinkage of eigenvalues toward their mean through a quadratic penalty on the likelihood is equivalent to assuming a normal distribution of the eigenvalues while the assumption of a double exponential

prior distribution results in a LASSO-type penalty (Huang *et al.* 2006).

Meyer and Kirkpatrick (2010) demonstrated that a REML analog to bending is obtained by imposing a penalty proportional to the variance of the canonical eigenvalues on the likelihood and showed that this can yield substantial reductions in loss for estimates of both  $\Sigma_G$  and  $\Sigma_E$ , the residual covariance matrix. Subsequent simulations (Meyer *et al.* 2011; Meyer 2011) examined the scope for penalties based on different functions of the parameters to be estimated and prior distributions for them and found them to be similarly effective, depending on the population values for the covariance matrices to be estimated.

A central component of the success of regularized estimation is the choice of how much to penalize. A common practice is to scale the penalty by a so-called “tuning factor” to regulate stringency of penalization. Various studies (again for a single covariance matrix, see above) demonstrated that this can be estimated reasonably well from the data at hand, using cross-validation techniques. Adopting these suggestions for genetic analyses and using  $k$ -fold cross-validation, Meyer (2011) estimated the appropriate tuning factor as that which maximized the average, unpenalized likelihood in the validation sets. However, this procedure was laborious and afflicted by problems in locating the maximum of a fairly flat likelihood surface for analyses involving many traits and not so large data sets. These technical difficulties all but prevented practical applications so far. Moreover, it was generally less successful than reported for studies considering a single covariance matrix. This led to the suggestion of imposing a mild penalty, determining the tuning factor as the largest value that did not cause a decrease in the (unpenalized) likelihood equivalent to a significant change in a single parameter. This pragmatic approach yielded reductions in loss that were generally of comparable magnitude to those achieved using cross-validation (Meyer 2011). However, it still required multiple analyses and thus considerably increased computational demands compared to standard, unpenalized estimation.

### Simple Penalties

In the Bayesian framework, the influence of the prior and thus the amount of regularization is generally specified through the so-called hyperparameters of the prior distribution, which determine its shape, scale, or location. This suggests that an alternative, tuning factor-free formulation for a penalty on the likelihood can be obtained by expressing it in terms of the distribution-specific (hyper)parameters. For instance, when assuming a normal prior for canonical eigenvalues, the regulating parameter is the variance of the normal distribution, with more shrinkage induced the lower its value. This may lend itself to applications employing default values for these parameters. Furthermore, such formulation may facilitate direct estimation of the regulating parameter, denoted henceforth as  $\nu$ , simultaneously with the covariance components to be estimated (G. de los Campos, personal

communication). In contrast, in a setting involving a tuning factor the penalized likelihood is, by definition, highest for a factor of zero (*i.e.*, no penalty) and thus does not provide this opportunity.

### Aims

This article examines the scope for REML estimation imposing penalties regulated by choosing the parameters of the selected prior distribution. Our aim is to obtain a formulation that allows uncomplicated use on a routine basis, free from laborious and complicated additional calculations, to determine the appropriate strength of penalization. It differs from previous work such that we do not strive to obtain maximum benefits, but are content with lesser—but still highly worthwhile—improvements in estimates that can be achieved through mild penalties, at little risk of detrimental effects for unusual cases where the population parameters do not match our assumptions for their prior distributions well. The focus is on penalties involving scale-free functions of covariance components that fall into defined intervals and are thus better suited to a choice of default-regulating parameters than functions that do not.

We begin with the description of suitable penalties together with a brief review of pertinent literature and outline the adaptation of standard REML algorithms. This is followed by a large-scale simulation study showing that the penalties proposed can yield substantial reductions in loss of estimates for a wide range of population parameters. In addition, the impact of penalized estimation is demonstrated for a practical data set, and the implementation of the methods proposed in our REML package *WOMBAT* is described. We conclude with a discussion and recommendations on selection of default parameters for routine use in multivariate analyses.

### Penalized Maximum-Likelihood Estimation

Consider a simple mixed, linear model for  $q$  traits with covariance matrices  $\Sigma_G$  and  $\Sigma_E$  due to additive genetic and residual effects, respectively, to be estimated. Let  $\log\mathcal{L}(\boldsymbol{\theta})$  denote the log-likelihood in a standard, unpenalized maximum-likelihood (or REML) analysis and  $\boldsymbol{\theta}$  denote the vector of parameters, composed of the distinct elements of  $\Sigma_G$  and  $\Sigma_E$  or the equivalent. The penalized likelihood is then (Green 1987)

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

with the penalty  $\mathcal{P}(\boldsymbol{\theta})$  a nonnegative function of the parameters to be estimated and  $\psi$  the so-called tuning factor that modulates the strength of penalization (the factor of 1/2 is used for algebraic consistency and could be omitted).

The penalty can be derived by assuming a suitable prior distribution for the parameters  $\boldsymbol{\theta}$  (or functions thereof) as minus the logarithmic value of the pertaining probability density. As in Bayesian estimation, the choice of the prior is often somewhat *ad hoc* or driven by aspects of convenience

(such as conjugacy of the priors) and computational feasibility. In the following, we assume  $\psi = 1$  throughout and regulate the amount of penalization instead via the parameters of the distribution from which the penalty is derived.

### Functions to be penalized

We consider two types of scale-free functions of the covariance matrices to be estimated as the basis for regularization.

**Canonical eigenvalues:** Following Hayes and Hill (1981), the first type comprises the canonical eigenvalues of  $\Sigma_G$  and  $\Sigma_P = \Sigma_G + \Sigma_E$ .

Multivariate theory shows that for two symmetric, positive definite matrices of the same size there is a transformation that yields  $\mathbf{T}\mathbf{T}' = \Sigma_P$  and  $\mathbf{T}\mathbf{A}\mathbf{T}' = \Sigma_G$ , with  $\mathbf{A}$  the diagonal matrix of canonical eigenvalues with elements  $\lambda_i$  (Anderson 1984). This can be thought of as transforming the traits considered to new variables that are uncorrelated and have phenotypic variance of unity; *i.e.*, the canonical eigenvalues are equal to heritabilities on the new scale and fall in the interval  $[0, 1]$  (Hayes and Hill 1980). It is well known that estimates of eigenvalues of covariance matrices are systematically overdispersed—the largest values are overestimated and the smallest are underestimated—while their mean is expected to be estimated correctly (Lawley 1956). Moreover, a major proportion of the sampling variation of covariance matrices can be attributed to this overdispersion of eigenvalues (Ledoit and Wolf 2004). Hence there have been various suggestions to modify the eigenvalues of sample covariance matrices in some way to reduce the loss in estimates; see Meyer and Kirkpatrick (2010) for a more detailed review.

**Correlations:** The second type of functions comprises correlations between traits, in particular genetic correlations. A number of Bayesian approaches to the estimation of covariance matrices decompose the problem into variances (or standard deviations) and correlations with separate priors, thus alleviating the inflexibility of the widely used conjugate prior given by an inverse Wishart distribution (Barnard *et al.* 2000; Daniels and Kass 2001; Zhang *et al.* 2006; Daniels and Pourahmadi 2009; Hsu *et al.* 2012; Bouriga and Féron 2013; Gaskins *et al.* 2014). However, overall few suitable families of prior density functions for correlation matrices have been considered and practical applications have been limited. In particular, estimation using Monte Carlo sampling schemes has been hampered by difficulties in sampling correlation matrices conforming to the constraints of positive definiteness and unit diagonals.

Most statistical literature concerned with Bayesian or penalized estimation of correlation matrices considered shrinkage toward an identity matrix, *i.e.*, shrinkage of individual correlations toward zero, although other, simple correlation structures have been proposed (Schäfer and Strimmer 2005). As outlined above, the motivation for bending (Hayes and Hill 1981) included the desire to borrow strength from the estimate of the phenotypic covariance matrix. Similar

arguments may support shrinkage of the genetic toward the phenotypic correlation matrix. This dovetails with what has become known as “Cheverud’s conjecture”: Reviewing a large body of literature, Cheverud (1988) found that estimates of genetic correlations were generally close to their phenotypic counterparts and thus proposed that phenotypic values should be substituted when genetic correlations could not be estimated. Subsequent studies reported similar findings for a range of traits in laboratory species, plants, and animals (e.g., Koots *et al.* 1994; Roff 1995; Waitt and Levin 1998).

### Partial correlations

Often a reparameterization can transform a constrained matrix problem to an unconstrained one. For instance, it is common practice in REML estimation of covariance matrices to estimate the elements of their Cholesky factors, coupled with a logarithmic transformation of the diagonal elements, to remove constraints on the parameter space (Meyer and Smith 1996). Pinheiro and Bates (1996) examined various transformations for covariance matrices and their impact on convergence behavior of maximum-likelihood analyses, and corresponding forms for correlation matrices have been described (Rapisarda *et al.* 2007).

To alleviate the problems encountered in sampling valid correlation matrices, Joe (2006) proposed a reparameterization of correlations to partial correlations, which vary independently over the interval  $[-1, 1]$ . This involves a one-to-one transformation between partial and standard correlations, *i.e.*, is readily reversible. Hence, using the reparameterization, we can readily sample a random correlation matrix that is positive definite by sampling individual, partial correlations. Moreover, it allows for flexible specification of priors in a Bayesian context by choosing individual and independent distributions for each one. Daniels and Pourahmadi (2009) referred to these quantities as partial autocorrelations (PACs), interpreting them as correlations between traits  $i$  and  $j$  conditional on the intervening traits,  $i + 1$  to  $j - 1$ .

**Details:** Consider a correlation matrix  $\mathbf{R}$  of size  $q \times q$  with elements  $\rho_{ij}$  (for  $i \neq j$ ) and  $\rho_{ii} = 1$ . As  $\mathbf{R}$  is symmetric, let  $i < j$ . For  $j = i + 1$ , the PACs are equal to the standard correlations,  $\pi_{i,i+1} = \rho_{i,i+1}$ , as there are no intervening variables. For  $j > i + 1$ , partition the submatrix of  $\mathbf{R}$  composed of rows and columns  $i$  to  $j$  as

$$\begin{pmatrix} 1 & \mathbf{r}'_1 & \rho_{ij} \\ \mathbf{r}_1 & \mathbf{R}_2 & \mathbf{r}_3 \\ \rho_{ij} & \mathbf{r}'_3 & 1 \end{pmatrix} \quad (2)$$

with  $\mathbf{r}_1$  and  $\mathbf{r}_3$  vectors of length  $j - i - 1$ , with elements  $\rho_{ik}$  and  $\rho_{jk}$ , respectively, and  $\mathbf{R}_2$  the corresponding matrix with elements  $\rho_{kl}$  for  $k, l = i + 1, \dots, j - 1$ . This gives PAC

$$\pi_{ij} = \frac{\rho_{ij} - \mathbf{r}'_1 \mathbf{R}_2^{-1} \mathbf{r}_3}{\sqrt{(1 - \mathbf{r}'_1 \mathbf{R}_2^{-1} \mathbf{r}_1)(1 - \mathbf{r}'_3 \mathbf{R}_2^{-1} \mathbf{r}_3)}} \quad (3)$$

and the reverse transformation is

$$\rho_{ij} = \mathbf{r}'_1 \mathbf{R}_2^{-1} \mathbf{r}_3 + \pi_{ij} \sqrt{(1 - \mathbf{r}'_1 \mathbf{R}_2^{-1} \mathbf{r}_1)(1 - \mathbf{r}'_3 \mathbf{R}_2^{-1} \mathbf{r}_3)} \quad (4)$$

(Joe 2006).

### Penalties

We derive penalties on canonical eigenvalues and correlations or partial autocorrelations, choosing independent Beta distributions as priors. This choice is motivated by the flexibility of this class of distributions and the property that its hyperparameters—which determine the strength of penalization—can be “set” by specifying a single value.

**Beta distribution:** The Beta distribution is a continuous probability function that is widely used in Bayesian analyses and encompasses functions with many different shapes, determined by two parameters,  $\alpha$  and  $\beta$ . While the standard Beta distribution is defined for the interval  $[0, 1]$ , it is readily extended to a different interval. The probability density function for a variable  $x \in [a, b]$  following a Beta distribution is of the form

$$p(x) = B(\alpha, \beta)^{-1} (b-a)^{-(\alpha+\beta-1)} (x-a)^{\alpha-1} (b-x)^{\beta-1} \quad (5)$$

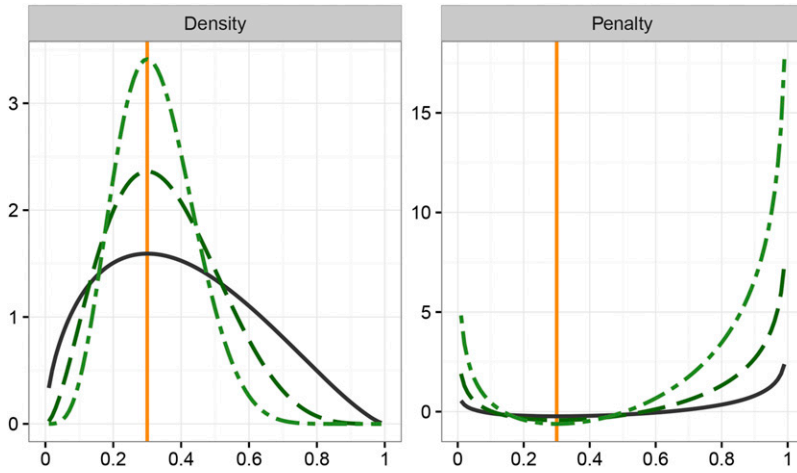
(Johnson *et al.* 1995, Chap. 25) with  $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$  and  $B(\cdot)$  and  $\Gamma - 4 \times (\cdot)$  denoting the Beta and Gamma functions, respectively.

When employing a Beta prior in Bayesian estimation, the sum of the shape parameters,  $\nu = \alpha + \beta$ , is commonly interpreted as the effective sample size of the prior (PSS) (Morita *et al.* 2008). It follows that we can specify the parameters of a Beta distribution with mode  $m$  as a function of the PSS ( $\nu \geq 2$ ) as

$$\alpha = 1 + \frac{m-a}{b-a}(\nu-2) \quad \text{and} \quad \beta = 1 + \frac{b-m}{b-a}(\nu-2). \quad (6)$$

For  $\nu > 2$ , this yields a unimodal distribution, and for  $m = (b-a)/2$  the distribution is symmetric, with  $\alpha = \beta$ . For given  $m$ , this provides a mechanism to regulate the strength of penalization through a single, intuitive parameter, the PSS  $\nu$ .

Figure 1 shows the probability density of a variable with a standard Beta distribution on the interval  $[0, 1]$  with mode of 0.3 together with the resulting penalty (*i.e.*, minus the logarithmic value of the density), for three values of PSS. For  $\nu = 2$ , the density function would be a uniform distribution, depicted by a horizontal line at height of 1, resulting in no penalty. With increasing values of  $\nu$ , the distribution becomes more and more peaked and the penalty on values close to the extremes of the range becomes more and more severe. Conversely, in spite of substantial differences in point mass around the mode, penalty values in proximity of the mode differ relatively little for different values of



**Figure 1** Probability densities (left) and corresponding penalties (including a factor of 1/2; right) for a variable with Beta distribution on  $[0, 1]$  with mode of  $m = 0.3$  for effective prior sample sizes of  $\nu = 4$  (—),  $\nu = 8$  (- -), and  $\nu = 16$  (- · -).

$\nu$ . While  $\mathcal{P}(\boldsymbol{\theta})$  in (1) was considered nonnegative, penalty values close to the mode can be negative—this does not affect suitability of the penalty and can be overcome by adding a suitable constant.

**Penalty on canonical eigenvalues:** Canonical eigenvalues fall in the interval  $[0, 1]$ . For  $q$  traits, there are likely to be  $q$  different values  $\lambda_i$  and attempts to determine the mode of their distribution may be futile. Hence we propose to substitute the mean canonical eigenvalue,  $\bar{\lambda}$ . Taking minus logarithmic values of (5) and assuming the same mode and PSS for all  $q$  eigenvalues then gives penalty

$$\mathcal{P}_\lambda = C - (\nu - 2) \left[ \bar{\lambda} \sum_{i=1}^q \log(\lambda_i) + (1 - \bar{\lambda}) \sum_{i=1}^q \log(1 - \lambda_i) \right] \quad (7)$$

with  $C = q \log \left[ B \left( 1 + \bar{\lambda}(\nu - 2), 1 + (1 - \bar{\lambda})(\nu - 2) \right) \right]$ . This formulation results in shrinkage of all eigenvalues toward  $\bar{\lambda}$ , with the highest and lowest values shrink the most.

**Penalty on correlations:** Following Joe (2006) and Daniels and Pourahmadi (2009), we assume independent shifted Beta distributions on  $[-1, 1]$  for PAC.

Daniels and Pourahmadi (2009) considered several Bayesian priors for correlation matrices formulated via PAC, suggesting uniform distributions for individual  $\pi_{ij}$ ; *i.e.*,  $\pi_{ij} \sim \text{Beta}(1, 1)$ . In addition, they showed that the equivalent to the joint uniform prior for  $\mathbf{R}$  proposed by Barnard *et al.* (2000),  $p(\mathbf{R}) \propto 1$ , is obtained by assuming Beta priors for PAC with shape parameters depending on the number of intervening variables; *i.e.*,  $\alpha_{i,i+k} = \beta_{i,i+k} = 1 + (q - 1 - k)/2$ . Similarly, priors proportional to higher powers of the determinant of  $\mathbf{R}$ ,  $p(\mathbf{R}) \propto |\mathbf{R}|^{t-1}$ , are obtained for  $\alpha_{i,i+k} = t + (q - 1 - k)/2$  (Daniels and Pourahmadi 2009). Gaskins *et al.* (2014) extended this framework to PAC-based priors with more aggressive shrinkage toward zero for higher lags, suitable to encourage sparsity in estimated correlation matrices for longitudinal data.

Both Joe (2006) and Daniels and Pourahmadi (2009) considered Beta priors for PAC with  $\alpha = \beta$ , *i.e.*, shrinkage of all  $\pi_{ij}$  toward zero. We generalize this by allowing for different shrinkage targets  $\tau_{ij}$ —and thus different shape parameters  $\alpha_{ij}$  and  $\beta_{ij}$ —for individual values  $\pi_{ij}$ . This gives penalty

$$\mathcal{P}_\pi = \frac{q(q-1)}{2} (\nu - 1) \log(2) + \sum_{i=1}^q \sum_{j=i+1}^q C_{ij} - \frac{\nu - 2}{2} \left[ (\tau_{ij} + 1) \log(\pi_{ij} + 1) + (1 - \tau_{ij}) \log(1 - \pi_{ij}) \right] \quad (8)$$

with  $C_{ij} = \log \left[ B \left( 1 + (\tau_{ij} + 1)(\nu - 2)/2, 1 + (1 - \tau_{ij})(\nu - 2)/2 \right) \right]$ . Again, this assumes equal PSS for all values, but could of course readily be expanded to allow for different values  $\nu_{ij}$  for different PACs. For all  $\tau_{ij} = 0$ , (8) reduces to

$$\mathcal{P}_\pi(0) = \frac{q(q-1)}{2} \left[ (\nu - 1) \log(2) + C_0 \right] - \frac{\nu - 2}{2} \sum_{i=1}^q \sum_{j=i+1}^q \log \left( 1 - \pi_{ij}^2 \right) \quad (9)$$

with  $C_0 = \log \left[ B \left( 1 + (\nu - 2)/2, 1 + (\nu - 2)/2 \right) \right]$ .

### Maximizing the penalized likelihood

REML estimation in quantitative genetics usually relies on algorithms exploiting derivatives of the log-likelihood function to locate its maximum. In particular, the so-called average information algorithm (Gilmour *et al.* 1995) is widely used due to its relative computational ease, good convergence properties, and implementation in several REML software packages. It can be described as a Newton(-Raphson)-type algorithm where the Hessian matrix is approximated by the average of observed and expected information. To adapt the standard, unpenalized algorithm for penalized estimation we need to adjust first and second derivatives of  $\log \mathcal{L}(\boldsymbol{\theta})$  for derivatives of the penalties with respect to the parameters,  $\theta_k$ , to be estimated. These differ if we choose fixed values to determine the modes of the assumed Beta priors (*e.g.*,  $\tau_{ij} = 0$

or the mean  $\lambda_i$  from a preliminary, unpenalized analysis) or employ penalties that derive these from the parameter estimates.

**Canonical eigenvalues:** If  $\bar{\lambda}$  in (7) is estimated from the data, its derivatives are nonzero. This gives

$$\begin{aligned} \frac{\partial \mathcal{P}_\lambda}{\partial \theta_k} &= \frac{\partial C}{\partial \theta_k} + (\nu - 2) \left\{ \frac{\partial \bar{\lambda}}{\partial \theta_k} \sum_{i=1}^q \log \left( \frac{1 - \lambda_i}{\lambda_i} \right) \right. \\ &\quad \left. + \sum_{i=1}^q \frac{\partial \lambda_i}{\partial \theta_k} \left( \frac{1 - \bar{\lambda}}{1 - \lambda_i} - \frac{\bar{\lambda}}{\lambda_i} \right) \right\} \end{aligned} \quad (10)$$

and

$$\begin{aligned} \frac{\partial^2 \mathcal{P}_\lambda}{\partial \theta_k \partial \theta_m} &= \frac{\partial^2 C}{\partial \theta_k \partial \theta_m} + (\nu - 2) \sum_{i=1}^q \frac{\partial^2 \bar{\lambda}}{\partial \theta_k \partial \theta_m} \log \left( \frac{1 - \lambda_i}{\lambda_i} \right) \\ &\quad - \left( \frac{\partial \bar{\lambda}}{\partial \theta_k} \frac{\partial \lambda_i}{\partial \theta_m} + \frac{\partial \lambda_i}{\partial \theta_k} \frac{\partial \bar{\lambda}}{\partial \theta_m} \right) \left( \frac{1}{\lambda_i} + \frac{1}{1 - \lambda_i} \right) \\ &\quad + \frac{\partial^2 \lambda_i}{\partial \theta_k \partial \theta_m} \left( \frac{1 - \bar{\lambda}}{1 - \lambda_i} - \frac{\bar{\lambda}}{\lambda_i} \right) \\ &\quad + \frac{\partial \lambda_i}{\partial \theta_k} \frac{\partial \lambda_i}{\partial \theta_m} \left( \frac{\bar{\lambda}}{\lambda_i^2} + \frac{1 - \bar{\lambda}}{(1 - \lambda_i)^2} \right). \end{aligned} \quad (11)$$

Derivatives of  $C$  involve the digamma and trigamma functions, e.g.,

$$\begin{aligned} \frac{\partial C}{\partial \theta_k} &= q(\nu - 2) \frac{\partial \bar{\lambda}}{\partial \theta_k} \left[ \psi(1 + (\nu - 2)\bar{\lambda}) \right. \\ &\quad \left. - \psi(1 + (\nu - 2)(1 - \bar{\lambda})) \right] \end{aligned}$$

with  $\psi$  the digamma function. Derivatives of  $\lambda_i$  required in (10) and (11) are easiest to evaluate if the analysis is parameterized to the canonical eigenvalues and the elements of the corresponding transformation matrix  $\mathbf{T}$  (see Meyer and Kirkpatrick 2010), so that  $\partial \lambda_i / \lambda_i = 1$  and  $\partial \bar{\lambda} / \lambda_i = 1/q$  and all other derivatives of  $\bar{\lambda}$  and  $\lambda_i$  are zero. A possible approximation is to ignore contributions of derivatives of  $\bar{\lambda}$ , arguing that the mean eigenvalue is expected to be unaffected by sampling overdispersion and thus should change little.

**Partial correlations:** Analogous arguments hold for penalties involving correlations. This gives

$$\begin{aligned} \frac{\partial \mathcal{P}_\pi}{\partial \theta_k} &= \sum_{i=1}^q \sum_{j=i+1}^q \frac{\partial C_{ij}}{\partial \theta_k} \\ &\quad + \frac{\nu - 2}{2} \left[ \log \left( \frac{1 - \pi_{ij}}{1 + \pi_{ij}} \right) \frac{\partial \tau_{ij}}{\partial \theta_k} + 2 \left( \frac{\pi_{ij} - \tau_{ij}}{1 - \pi_{ij}^2} \right) \frac{\partial \pi_{ij}}{\partial \theta_k} \right] \end{aligned} \quad (12)$$

and

$$\begin{aligned} \frac{\partial^2 \mathcal{P}_\pi}{\partial \theta_k \partial \theta_m} &= \sum_{i=1}^q \sum_{j=i+1}^q \frac{\partial^2 C_{ij}}{\partial \theta_k \partial \theta_m} + \frac{\nu - 2}{2} \left[ \log \left( \frac{1 - \pi_{ij}}{1 + \pi_{ij}} \right) \frac{\partial^2 \tau_{ij}}{\partial \theta_k \partial \theta_m} \right. \\ &\quad - \frac{2}{1 - \pi_{ij}^2} \left( \frac{\partial \tau_{ij}}{\partial \theta_k} \frac{\partial \pi_{ij}}{\partial \theta_m} + \frac{\partial \tau_{ij}}{\partial \theta_m} \frac{\partial \pi_{ij}}{\partial \theta_k} \right) \\ &\quad + 2 \left( \frac{\pi_{ij} - \tau_{ij}}{1 - \pi_{ij}^2} \right) \frac{\partial^2 \pi_{ij}}{\partial \theta_k \partial \theta_m} \\ &\quad \left. - 2 \left( \frac{1 - \tau_{ij}^2 + (\pi_{ij} - \tau_{ij})^2}{(1 - \pi_{ij}^2)^2} \right) \frac{\partial \pi_{ij}}{\partial \theta_k} \frac{\partial \pi_{ij}}{\partial \theta_m} \right] \end{aligned} \quad (13)$$

with obvious simplifications if shrinkage targets are fixed or treated as such, so that derivatives of  $\tau_{ij}$  are zero. As shown in the *Appendix*, derivatives of correlations and PACs are readily calculated from the derivatives of covariance components for any of the parameterizations commonly utilized in (unpenalized) REML algorithms for variance component estimation.

## Simulation Study

A large-scale simulation study, considering a wide range of population parameters, was carried out to examine the efficacy of the penalties proposed above.

### Setup

Data were sampled from multivariate normal distributions for  $q = 9$  traits, assuming a balanced paternal half-sib design composed of  $s$  unrelated sire families with 10 progeny each. Sample sizes considered were  $s = 100, 400, \text{ and } 1000$ , with records for all traits for each of the progeny but no records for sires.

Population values for genetic and residual variance components were generated by combining 13 sets of heritabilities with six different types of correlation structures to generate 78 cases. Details are summarized in the *Appendix* and population canonical eigenvalues for all sets are shown in Supplemental Material, [File S1](#). To assess the potential for detrimental effects of penalized estimation, values were chosen deliberately to generate both cases that approximately matched the priors assumed in deriving the penalties and cases where this was clearly not the case. The latter included scenarios where population canonical eigenvalues were widely spread and in multiple clusters and cases where genetic and phenotypic correlations were highly dissimilar. A total of 500 samples per case and sample size were obtained and analyzed.

### Analyses

REML estimates of  $\Sigma_G$  and  $\Sigma_E$  for each sample were obtained without penalization and imposing a penalty on canonical eigenvalues  $\mathcal{P}_\lambda$ , as given in (7), and penalties on partial autocorrelations shrinking all values toward zero [ $\mathcal{P}_\pi(0)$ , see (9)] and with shrinkage targets for each value equal to the corresponding phenotypic counterpart [ $\mathcal{P}_\pi(P)$ , see (8)]. For



the latter, penalties on genetic PAC only and both genetic and residual values were examined.

Analyses were carried out considering fixed values for the effective sample size, ranging from  $\nu = 2.5$  to 24. For penalties on both genetic and residual PAC, either the same value was used for both or the PSS for residual PAC was fixed at  $\nu = 8$ . In addition, direct estimation of a suitable PSS for each replicate was attempted. As shown in (1), penalties were subtracted from the standard log-likelihood, incorporating a factor of  $1/2$ .

The model of analysis was a simple animal model, fitting means for each trait as the only fixed effects. A method-of-scoring algorithm together with simple derivative-free search steps was used to locate the maximum of the (penalized) likelihood function. To facilitate easy computation of derivatives of  $\mathcal{P}_\lambda$ , this was done using a parameterization to the elements of the canonical decomposition (see Meyer and Kirkpatrick 2010), restraining estimates of  $\lambda_i$  to the interval of  $[0.0001, 0.9999]$ .

Direct estimation of  $\nu$  was performed by evaluating points on the profile likelihood for  $\nu$  [*i.e.*, maximizing  $\log \mathcal{L}_P(\boldsymbol{\theta})$  with respect to the covariance components to be estimated for selected, fixed values of  $\nu$ ], combined with quadratic approximation steps of the profile to locate its maximum, using Powell's (2006) Fortran subroutine NEWUOA. To avoid numerical problems, estimates of  $\nu$  were constrained to the interval  $[2.01, 50]$ . All calculations were carried out using custom Fortran programs (available on request).

### Summary statistics

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

$$L_1(\boldsymbol{\Sigma}_X, \hat{\boldsymbol{\Sigma}}_X) = \text{tr}(\boldsymbol{\Sigma}_X^{-1} \hat{\boldsymbol{\Sigma}}_X) - \log \left| \boldsymbol{\Sigma}_X^{-1} \hat{\boldsymbol{\Sigma}}_X \right| - q \quad (14)$$

for  $X = G, E$ , and  $P$ , with  $\boldsymbol{\Sigma}_X$  denoting the matrix of population values for genetic, residual, and phenotypic covariances and  $\hat{\boldsymbol{\Sigma}}_X$  the corresponding estimate. As suggested by Lin and Perlman (1985), the percentage reduction in average loss (PRIAL) was used as the criterion to summarize the effects of penalization

$$100 \left[ 1 - \bar{L}_1(\boldsymbol{\Sigma}_X, \hat{\boldsymbol{\Sigma}}_X^\nu) / \bar{L}_1(\boldsymbol{\Sigma}_X, \hat{\boldsymbol{\Sigma}}_X^0) \right], \quad (15)$$

where  $\bar{L}_1(\cdot)$  denotes the entropy loss averaged over replicates, and  $\hat{\boldsymbol{\Sigma}}_X^\nu$  and  $\hat{\boldsymbol{\Sigma}}_X^0$  represent the penalized and corresponding unpenalized REML estimates of  $\boldsymbol{\Sigma}_X$ , respectively.

In addition, the average reduction in  $\log \mathcal{L}(\boldsymbol{\theta})$  due to penalization,  $\Delta \mathcal{L}$ , was calculated as the mean difference across replicates between the unpenalized likelihood for estimates  $\hat{\boldsymbol{\Sigma}}_X^\nu$  and the corresponding value for estimates  $\hat{\boldsymbol{\Sigma}}_X^0$ .

### Data availability

The author states that all data necessary for confirming the conclusions presented in the article are represented fully within the article.

## Results

Distributions of PRIAL across the 78 sets of population values together with the corresponding change in likelihood are shown in Figure 2 for two sample sizes, with penalties  $\mathcal{P}_\pi(0)$  and  $\mathcal{P}_\pi(P)$  applied to genetic PAC only. Distributions shown are trimmed; *i.e.*, their range reflects minimum and maximum values observed. Selected mean and minimum values are also reported in Table 1. More detailed results for individual cases are reported in File S1.

### Genetic covariances

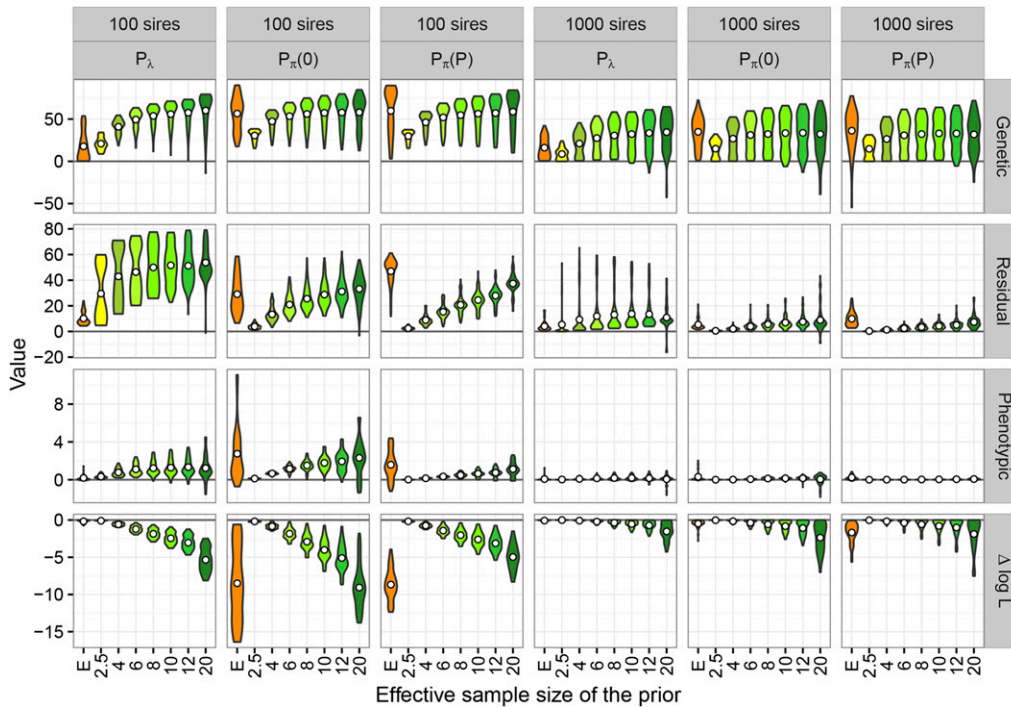
Overall, for fixed PSS there were surprisingly few differences between penalties in mean PRIAL values achieved, especially for estimates of the genetic covariance matrix. Correlations between PRIAL for  $\hat{\boldsymbol{\Sigma}}_G$  from 0.9 to unity suggested similar modes of action.

For comparison, additional analyses considered penalties on standard correlations, obtained by substituting  $\rho_{ij}$  for  $\pi_{ij}$  in (8), (9), (12), and (13). Corresponding PRIAL values (not shown) for these were consistently lower and, more importantly, a considerable number of minimum values were negative, even for small values of the PSS. Clearly this reflected a discrepancy between priors assumed and the distribution of correlations. Transformation to partial autocorrelation yielded a better match and thus yielded penalties markedly less likely to have detrimental effects. While easier to interpret than PAC, penalties on standard correlations based on independent Beta priors are thus not recommended.

Even for small values of  $\nu$  there were worthwhile reductions in loss for estimates of  $\boldsymbol{\Sigma}_G$ , in particular for the smallest sample ( $s = 100$ ). Means increased with increasing stringency of penalization along with an increasing spread in results for individual cases, especially for the largest sample size. This pattern was due to the range of population values for genetic parameters used.

Moreover, for small samples or low PSS, it did not appear to be all that important whether the priors on which the penalties were based approximately matched population values or not: "Any" penalty proved beneficial, *i.e.*, resulted in positive PRIAL for  $\hat{\boldsymbol{\Sigma}}_G$ . For more stringent penalization, however, there was little improvement (or even adverse effects) for the cases where there was a clear mismatch. For instance, for  $\mathcal{P}_\lambda$  for  $\nu = 24$  and  $s = 100$  sires, there were two cases with negative PRIAL for  $\hat{\boldsymbol{\Sigma}}_G$ . Both of these had a cluster of high and low population values for  $\lambda_i$  so that the assumption of a unimodal distribution invoked in deriving  $\mathcal{P}_\lambda$  was inappropriate and led to sufficient overshrinkage to be detrimental. On the whole, however, unfavorable effects of penalization were few and restricted to the most extreme cases considered.

Paradoxically, PRIAL values for  $\hat{\boldsymbol{\Sigma}}_G$  were also relatively low for cases where heritabilities were approximately the same and genetic and phenotypic correlations were similar, so that canonical eigenvalues differed little from their mean (see File S1). This could be attributed to the shape of the



**Figure 2** Distribution of percent reduction in average loss for estimates of genetic, residual, and phenotypic covariance matrices, together with corresponding change in log-likelihood ( $\Delta\mathcal{L}$ ) for penalties on canonical eigenvalues ( $\mathcal{P}_\lambda$ ) and genetic, partial autocorrelations, shrinking toward zero [ $\mathcal{P}_\pi(0)$ ] or phenotypic values [ $\mathcal{P}_\pi(P)$ ]. Centered circles give mean values. Numerical values on the x-axis are fixed, prior effective sample sizes while “E” denotes the use of a value estimated from the data for each replicate.

penalty function, as illustrated in Figure 1, resulting in little penalization for values close to the mode. In other words, these were cases where the prior also did not quite match the population values: While the assumption of a common mean for canonical eigenvalues clearly held, that of a distribution on the interval  $[0, 1]$  did not. This can be rectified by specifying a more appropriate interval. As the unpenalized estimates of  $\lambda_i$  are expected to be overdispersed, their extremes may provide a suitable range to be used. Additional simulations (not shown) for  $\mathcal{P}_\lambda$  replaced values of  $a = 0$  and  $b = 1$  used to derive  $\mathcal{P}_\lambda$  in (7) with  $a = \max(0, \hat{\lambda}_9 - 0.05)$  and  $b = \min(1, \hat{\lambda}_1 + 0.05)$  for each replicate, where  $\hat{\lambda}_1$  and  $\hat{\lambda}_9$  represented the largest and smallest canonical eigenvalue estimate from a preliminary, unpenalized analysis, respectively. This increased PRIAL for both  $\hat{\Sigma}_G$  and  $\hat{\Sigma}_E$  substantially for these cases. However, as the proportion of such cases was low (see File S1), overall results were little affected.

### Residual covariances

Differences between penalties were apparent for  $\hat{\Sigma}_E$ .  $\mathcal{P}_\lambda$  involved terms  $\log(1 - \lambda_i)$  [see (7)], *i.e.*, the canonical eigenvalues of  $\hat{\Sigma}_E$  and  $\hat{\Sigma}_P$ . Hence,  $\mathcal{P}_\lambda$  yielded substantial PRIAL for estimates of both  $\hat{\Sigma}_G$  and  $\hat{\Sigma}_E$ , especially for the smaller samples where sampling variances and losses were high. Conversely, applying penalties on genetic PAC resulted in some, but lower improvements in  $\hat{\Sigma}_E$  (except for large  $\nu$ ), but only as a by-product due to negative sampling correlations between  $\hat{\Sigma}_G$  and  $\hat{\Sigma}_E$ . As shown in Table 2, imposing a corresponding penalty on residual PACs in addition could increase the PRIAL in estimates of  $\hat{\Sigma}_E$  markedly without reduction in the PRIAL for  $\hat{\Sigma}_G$ , provided the PSS chosen corresponded to a relatively mild degree of penalization. Shrinking toward

phenotypic PAC yielded somewhat less spread in PRIAL for  $\hat{\Sigma}_E$  than shrinking toward zero, accompanied by smaller changes in  $\log\mathcal{L}(\theta)$ .

### Phenotypic covariances

We argued above that imposing penalties based on the estimate of  $\hat{\Sigma}_P$  would allow us to borrow strength from it because typically  $\hat{\Sigma}_P$  is estimated more precisely than any of its components. Doing so, we would hope to have little—and certainly no detrimental—effect on the estimates of  $\hat{\Sigma}_P$ . Loosely speaking, we would expect penalized estimation to redress, to some extent at least, any distortion in partitioning of  $\hat{\Sigma}_P$  due to sampling correlations. As demonstrated in Figure 2, this was generally the case for fixed values of the PSS less than  $\sim\nu = 10$  or 12. Conversely, negative PRIAL for estimates of  $\hat{\Sigma}_P$  for higher values of  $\nu$  flagged overpenalization for cases where population values for genetic parameters did not sufficiently match the assumptions on which the penalties were based.

### Canonical eigenvalues

Figure 3 shows the distribution of the largest and smallest canonical eigenvalues, contrasting population values with mean estimates from unpenalized and penalized analyses for a medium sample size and a fixed PSS of  $\nu = 8$ . Results clearly illustrate the upward bias in estimates of the largest and downward bias in estimates of the smallest eigenvalues. As expected, imposing penalty  $\mathcal{P}_\lambda$  reduced the mean of the largest and increased the mean of the smallest eigenvalues, with some overshrinkage, especially of the largest eigenvalue, evident. In contrast, for the small value of  $\nu = 8$  chosen, the distribution of the largest values from penalized and



**Table 1** Selected mean and minimum values for percentage reduction in average loss for estimates of genetic ( $\Sigma_G$ ), residual ( $\Sigma_E$ ), and phenotypic ( $\Sigma_P$ ) covariance matrices together with mean change in unpenalized log-likelihood from the maximum ( $\Delta\mathcal{L}$ ) for penalties on canonical eigenvalues ( $\mathcal{P}_\lambda$ ) and genetic correlations, shrinking partial autocorrelations toward zero [ $\mathcal{P}_\pi(0)$ ] or phenotypic values [ $\mathcal{P}_\pi(P)$ ]

Penalty	$\nu^a$	100 sires				400 sires				1000 sires			
		$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$
Mean values													
$\mathcal{P}_\lambda$	4	41	43	1	-0.6	31	19	0	-0.2	21	9	0	-0.1
	6	49	46	1	-1.2	38	22	0	-0.5	27	12	0	-0.2
	8	54	50	1	-1.9	42	24	0	-0.7	30	13	0	-0.4
	12	58	51	1	-3.0	45	25	0	-1.4	33	13	0	-0.7
	<i>E</i>	18	10	0	-0.2	22	7	0	-0.1	16	4	0	-0.1
$\mathcal{P}_\pi(0)$	4	47	13	1	-0.9	37	5	0	-0.3	27	2	0	-0.2
	6	53	21	1	-1.8	43	9	0	-0.7	31	4	0	-0.4
	8	56	26	2	-2.9	44	12	0	-1.1	32	6	0	-0.6
	12	58	31	2	-5.1	45	16	1	-2.1	34	8	0	-1.1
	<i>E</i>	56	29	3	-8.5	46	12	1	-1.9	35	5	0	-0.5
$\mathcal{P}_\pi(P)$	4	46	9	0	-0.7	37	3	0	-0.3	26	1	0	-0.2
	6	52	15	0	-1.4	42	6	0	-0.7	31	2	0	-0.4
	8	55	21	0	-2.0	44	8	0	-1.0	32	3	0	-0.6
	12	57	28	1	-3.1	45	11	0	-1.6	33	5	0	-1.0
	<i>E</i>	60	47	2	-8.7	50	22	1	-3.5	36	10	0	-1.7
Minimum values													
$\mathcal{P}_\lambda$	8	12	26	0	-2.9	5	7	0	-1.7	2	2	0	-1.1
	12	1	13	0	-4.7	-11	11	-1	-3.0	-14	4	-1	-2.2
	<i>E</i>	0	4	0	-0.7	1	1	0	-0.6	1	0	0	-0.4
$\mathcal{P}_\pi(0)$	8	16	11	0	-5.0	2	2	0	-3.0	1	0	0	-1.9
	12	14	11	0	-8.7	-7	2	-1	-5.4	-13	0	-1	-3.4
	<i>E</i>	18	7	-1	-16.4	1	3	-1	-11.6	1	1	0	-2.8
$\mathcal{P}_\pi(P)$	8	18	7	0	-3.5	4	1	0	-2.9	1	0	0	-2.6
	12	16	12	0	-5.4	-1	2	0	-4.7	-5	0	0	-4.3
	<i>E</i>	3	12	-1	-12.4	-34	7	0	-8.8	-54	3	0	-5.6

<sup>a</sup>Effective sample size of the Beta prior. Numerical values are "fixed" values used while *E* denotes values estimated from the data for each replicate, with means of 3.1, 5.9, and 8.5 (for 100, 400, and 1000 sires, respectively) for  $\mathcal{P}_\lambda$ ; 31.2, 16.0, and 13.9 for  $\mathcal{P}_\pi(0)$ ; 45.4, 36.4, and 34.8 for  $\mathcal{P}_\pi(P)$ ; and 34.0, 33.6, and 32.2 for  $\mathcal{P}_\rho(P)$ .

unpenalized analyses differed little for penalty  $\mathcal{P}_\pi(P)$ ; *i.e.*, penalizing genetic PAC did not affect the leading canonical eigenvalues markedly, acting predominately on the smaller values. For more stringent penalties, however, some shrinkage of the leading eigenvalues due to penalties  $\mathcal{P}_\pi(P)$  and  $\mathcal{P}_\pi(0)$  was evident; detailed results for selected cases are given in File S1.

### Estimating PSS

On the whole, attempts to estimate the appropriate value of  $\nu$  from the data were not all that successful. For  $\mathcal{P}_\lambda$ , numerous cases yielded an estimate of  $\nu$  close to the lower end of the range allowed, *i.e.*, virtually no penalty. Conversely, for  $\mathcal{P}_\pi(0)$  and  $\mathcal{P}_\pi(P)$  a substantial number of cases resulted in estimates of  $\nu$  close to the upper bound allowed. This increased PRIAL (compared to fixed values for  $\nu$ ) for cases that approximately matched the priors but caused reduced or negative PRIAL and substantial changes in  $\log\mathcal{L}(\theta)$  otherwise. A possible explanation was that the penalized likelihood and thus the estimate of  $\nu$  might be dominated by  $\sum_{E^*}$ . However, as shown in Table 2, neither estimating a value for  $\Sigma_G$  ( $\nu_G$ ) while fixing the PSS for  $\Sigma_E$  ( $\nu_E$ ) nor estimating a value for both (either separately or jointly,  $\nu_G = \nu_E$ ) improved results greatly. Moreover, it yielded more cases for which penalization resulted in substantial, negative PRIAL, especially for  $\mathcal{P}_\pi(P)$ . Repeating

selected analyses using a sequential grid search to determine optimal values of  $\nu$  gave essentially the same results; *i.e.*, results could not be attributed to inadequacies in the quadratic approximation procedure. Given the additional computational requirements and the fact that suitable values could not be estimated reliably, joint estimation of PSS together with the covariance components cannot be recommended.

### Discussion

Sampling variation is the bane of multivariate analyses in quantitative genetics. While nothing can replace large numbers of observations with informative data and appropriate relationship structure, we often need to obtain reasonably trustworthy estimates of genetic parameters from relatively small data sets. This holds especially for data from natural populations but is also relevant for new or expensive to measure traits in livestock improvement or plant breeding schemes. We have shown that regularized estimation in a maximum-likelihood framework through penalization of the likelihood function can provide "better" estimates of covariance components, *i.e.*, estimates that are closer to the population values than those from standard, unpenalized analyses. This is achieved through penalties targeted at reducing sampling variation.

**Table 2** Selected mean and minimum values for percentage reduction in average loss for estimates of genetic ( $\Sigma_G$ ), residual ( $\Sigma_E$ ), and phenotypic ( $\Sigma_P$ ) covariance matrices together with mean change in unpenalized log-likelihood from the maximum ( $\Delta\mathcal{L}$ ) for penalties on both genetic and residual partial autocorrelations, shrinking toward zero [ $\mathcal{P}_\pi(0)$ ] or phenotypic values [ $\mathcal{P}_\pi(P)$ ]

Penalty	$\nu^a$		100 sires				400 sires				1000 sires				
	$\nu_G$	$\nu_E$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	$\Sigma_G$	$\Sigma_E$	$\Sigma_P$	$\Delta\mathcal{L}$	
Mean values															
$\mathcal{P}_\pi(0)$	4	4	48	38	1	-0.9	37	16	0	-0.4	27	7	0	-0.2	
	8	4	56	43	2	-2.9	45	20	0	-1.1	33	10	0	-0.6	
	8	8	57	48	2	-2.8	45	23	1	-1.1	33	12	0	-0.6	
	<i>E</i>	4	57	46	3	-6.6	48	21	1	-1.3	35	10	0	-0.5	
	<i>E</i>	$\nu_G$	61	55	3	-5.3	50	27	1	-1.8	37	14	0	-0.8	
$\mathcal{P}_\pi(P)$	<i>E</i>	<i>E</i>	53	43	3	-9.1	47	24	1	-1.9	35	13	0	-0.5	
	4	4	47	37	0	-0.8	37	15	0	-0.4	27	7	0	-0.2	
	8	4	55	43	1	-2.2	44	19	0	-1.0	32	9	0	-0.6	
	8	8	56	47	1	-2.3	44	22	0	-1.1	32	11	0	-0.6	
	<i>E</i>	4	60	57	1	-8.2	51	31	1	-3.5	36	14	0	-1.6	
$\mathcal{P}_\pi(0)$	<i>E</i>	$\nu_G$	61	60	2	-10.8	46	31	1	-6.5	33	18	0	-3.0	
	<i>E</i>	<i>E</i>	61	60	2	-10.7	51	33	1	-4.9	36	15	0	-2.3	
	Minimum values														
	$\mathcal{P}_\pi(0)$	4	4	17	11	0	-1.5	2	1	0	-0.9	0	0	0	-0.6
		8	4	17	17	1	-4.8	3	3	0	-2.9	1	1	0	-1.9
8		8	16	21	0	-4.6	3	4	0	-2.9	1	1	0	-2.0	
<i>E</i>		4	16	11	0	-12.8	5	4	0	-5.9	1	2	0	-2.9	
<i>E</i>		$\nu_G$	17	21	-1	-10.6	3	6	0	-6.2	2	2	0	-3.0	
$\mathcal{P}_\pi(P)$	<i>E</i>	<i>E</i>	-70	-139	-5	-17.7	6	-69	-2	-12.3	2	-5	0	-2.9	
	4	4	20	10	0	-1.3	2	1	0	-1.0	0	0	0	-0.8	
	8	4	21	20	0	-3.7	4	3	0	-3.0	1	1	0	-2.7	
	8	8	20	23	0	-3.8	5	5	0	-3.1	1	1	0	-2.7	
	<i>E</i>	4	5	24	-1	-10.8	-28	9	0	-7.9	-48	3	0	-4.9	
$\mathcal{P}_\pi(P)$	<i>E</i>	$\nu_G$	-3	10	-1	-17.0	-32	-28	-3	-16.0	-65	1	-1	-8.5	
	<i>E</i>	<i>E</i>	-1	10	-2	-17.4	-30	-6	-2	-9.4	-64	-18	-2	-6.7	

<sup>a</sup>Effective sample size of the Beta prior for genetic ( $\nu_G$ ) and residual ( $\nu_E$ ) correlations. Numbers given are fixed values used while *E* denotes values estimated from the data for each replicate.

### Simplicity

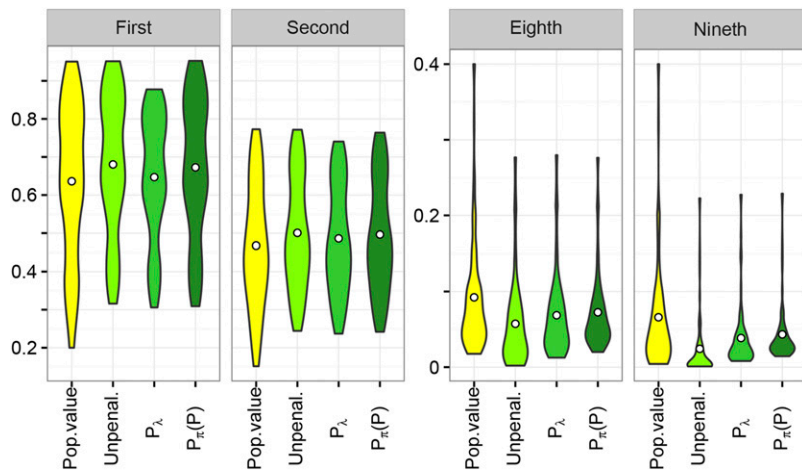
The aim of this study was to obtain a procedure to improve REML estimates of covariance components and thus genetic parameters that is easy to use without considerable increase in computational requirements and suitable for routine application. We have demonstrated that it is feasible to choose default values for the strength of penalization that yield worthwhile reductions in loss for a wide range of scenarios and are robust, *i.e.*, are unlikely to result in penalties with detrimental effects, and are technically simple. While such a tuning-free approach may not yield a maximum reduction in loss, it appears to achieve a substantial proportion thereof in most cases with modest changes in the likelihood compared to the maximum of  $\log\mathcal{L}(\theta)$  (without penalization). In contrast to earlier attempts to estimate a tuning factor (Meyer and Kirkpatrick 2010; Meyer 2011), it does not require multiple additional analyses to be carried out, and effects of penalization on computational requirements are thus mostly unimportant.

In addition, we can again make the link to Bayesian estimation, where the idea of mildly or weakly informative priors has been gaining popularity. Discussing priors for variance components in hierarchical models, Gelman (2006) advocated a half-t or half-Cauchy prior with large-scale

parameters. Huang and Wand (2013) extended this to prior distributions for covariance matrices that resulted in half-t priors for standard deviations and marginal densities of correlations  $\rho$  proportional to a power of  $(1 - \rho^2)$ . Chung *et al.* (2015) proposed a prior for covariance matrices proportional to a Wishart distribution with a diagonal scale matrix and low degrees of belief to obtain a penalty on the likelihood function that ensured nondegenerate estimates of variance components.

### Choice of penalty

We have presented two types of suitable penalties that fit well within the standard framework of REML estimation. Both achieved overall comparable reductions in loss but acted slightly differently, with penalties on correlations mainly affecting the smallest eigenvalues of the covariance matrices while penalties on canonical eigenvalues acted on both the smallest and largest values. Clearly it is the effect on the smallest eigenvalues (which have the largest sampling variances) that contributes most to the overall reduction in loss for a covariance matrix. An advantage of the penalty on correlations is that it is readily implemented for the parameterizations commonly employed in REML estimation, and it is straightforward to extend it to models with additional random effects and covariance matrices to be estimated or cases where



**Figure 3** Distribution of mean estimates of selected canonical eigenvalues comparing results from unpenalized analyses and analyses imposing penalties on canonical eigenvalues ( $P_\lambda$ ) and genetic, partial autocorrelations, shrinking toward phenotypic values [ $P_\pi(P)$ ], with population values in the simulation. Both penalties employ a fixed, prior effective sample size of  $\nu = 8$ ; results shown are for samples with 400 sire families. Centered circles give mean values across the 78 cases considered.

traits are recorded on distinct subsets of individuals so that some residual covariances are zero. It also lends itself to scenarios where we may be less interested in a reduction in sampling variance but may want to shrink correlations toward selected target values.

### Strength of penalization

Results suggest that penalties on canonical eigenvalues or PAC assuming a Beta prior with a conservative choice of PSS of  $\nu = 4-10$  will not only result in substantial improvements in estimates of genetic covariance components for many cases, but also result in little chance of detrimental effects for cases where the assumed prior does not quite match the underlying population values. Reanalyzing a collection of published heritability estimates from Mousseau and Roff (1987), M. Kirkpatrick (personal communication) suggested that their empirical distribution could be modeled as  $\text{Beta}(1.14, 1.32)$ , corresponding to  $\nu = 2.46$  and mode of 0.3.

### Additional bias

The price for a reduction in sampling variances from penalized analyses is an increase in bias. It is often forgotten that REML estimates of covariance components are biased even if no penalty is applied, as estimates are constrained to the parameter space; *i.e.*, the smallest eigenvalues are truncated at zero or, in practice, at a small positive value to ensure estimated matrices are positive definite. As shown in Figure 3, penalization tended to increase the lower limits for the smallest canonical eigenvalues and thus also for the corresponding values of the genetic covariance matrix, thus adding to the inherent bias. Previous work examined the bias due to penalization on specific genetic parameters in more detail (Meyer and Kirkpatrick 2010; Meyer 2011), showing that changes from unpenalized estimates were usually well within the range of standard errors. Employing a mild penalty with fixed PSS, changes in  $\log\mathcal{L}(\theta)$  from the maximum value in an unpenalized analysis were generally small and well below significance levels. This suggests that additional bias in REML estimates due to a mild penalty on the likelihood is of minor

concern and far outweighed by the benefits of a reduction in sampling variance.

### Alternatives

Obviously, there are many other choices. As in Bayesian estimation, results of penalized estimation are governed by the prior distribution selected, the more so the smaller the data set available. Mean reductions in loss obtained in a previous study, attempting to estimate tuning factors and using penalties derived assuming a normal distribution of canonical eigenvalues or inverse Wishart distributions of covariance or correlation matrices, again were by and large of similar magnitude (Meyer 2011). Other opportunities to reduce sampling variation arise through more parsimonious modeling, *e.g.*, by estimating  $\Sigma_G$  at reduced rank or assuming a factor-analytic structure. Future work should examine the scope for penalization in this context and consider the effects on model selection.

### Application

Practical application of REML estimation using the simple penalties proposed is illustrated using the data set considered by Meyer and Kirkpatrick (2010) (available as supplemental material in their article). In brief, this was composed of six traits measured on beef cattle, with 912–1796 records per trait. Including pedigree information on parents, grandparents, etc., without records yielded a total of 4901 animals in the analysis. The model of analysis fitted animals' additive genetic effects as random and so-called contemporary group effects, with up to 282 levels per trait, as fixed. The latter was a subclass representing a combination of systematic environmental factors to ensure only animals subject to similar conditions were directly compared.

Estimates of  $\Sigma_G$  and  $\Sigma_E$  were obtained by REML using WOMBAT (Meyer 2007), applying penalties on canonical eigenvalues and partial correlations (shrinking toward zero) using PSS of 4 and 8, and contrasted to unpenalized results. Lower bound sampling errors for estimates of canonical

eigenvalues were approximated empirically (Meyer and Houle 2013).

Results are summarized in Table 3. For all penalties and both levels of PSS, deviations in the unpenalized likelihood from the maximum (for the “standard” unpenalized analysis) were small and well below the significance threshold of  $-1.92$  for a likelihood-ratio test involving a single parameter (at an error probability of 5%), emphasizing the mildness of penalization even for the larger PSS. Similarly, changes in estimates of canonical eigenvalues were well within their 95% confidence limits, except for  $\lambda_1$  for  $\mathcal{P}_\lambda$  with  $\nu = 8$ , which was just outside the lower limit. As expected from simulation results above (cf. Figure 3), the penalty on canonical eigenvalues affected estimates of the largest values more than penalties on partial correlations. Conversely, the latter yielded somewhat larger changes in the lowest eigenvalues. Estimates for  $\mathcal{P}_\lambda$  with  $\nu = 8$  agreed well with results from Meyer and Kirkpatrick (2010) (of 0.69, 0.50, 0.38, 0.27, 0.17, and 0.05 for  $\lambda_1$ – $\lambda_6$ , respectively), which were obtained by applying a quadratic penalty on the canonical eigenvalues and using cross-validation to determine the stringency of penalization.

Corresponding changes in individual parameters were small throughout and again well within the range of  $\sim 95\%$  confidence intervals (sampling errors derived from the inverse of the average information matrix) for the unpenalized analysis. The largest changes in correlations occurred for the pairs of traits with the smallest numbers of records and the higher PSS. For instance, for  $\mathcal{P}_\lambda$  the residual correlation between traits 1 and 4 changed from 0.82 (with standard error of 0.17) to 0.65, while  $\mathcal{P}_\pi(0)$  on both genetic and residual PAC reduced the estimate of the genetic correlation from 0.48 (standard error of 0.25) to 0.37. Plots with detailed results for individual parameters are given in File S1.

In summary, none of the penalties applied changed estimates significantly and none of the changes in estimates of individual parameters due to penalization were questionable. Indeed, the larger changes described yielded values more in line with literature results for the traits concerned, suggesting that penalization stabilized somewhat erratic estimates based on small numbers of observations.

## Implementation

Penalized estimation for the penalties proposed for fixed values of  $\nu$  has been implemented in our mixed-model package WOMBAT (available at <http://didgeridoo.une.edu.au/km/wombat.php>) (MEYER 2007). For  $\mathcal{P}_\lambda$  a parameterization to the elements of the canonical decomposition is used for ease of implementation, while penalties on correlations use the standard parameterization to elements of the Cholesky factors of the covariance matrices to be estimated. Maximization of the likelihood is carried out using the average information algorithm, combined with derivative-free search steps where necessary to ensure convergence. Example runs for a simulated data set are shown in File S1.

**Table 3 Reduction in unpenalized likelihood ( $\Delta\mathcal{L}$ ) for applied example together with estimates of canonical eigenvalues**

Penalty <sup>a</sup>	$\nu^b$	$\Delta\mathcal{L}$	Canonical eigenvalues					
			$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$
None <sup>c</sup>		0	0.887	0.543	0.383	0.239	0.138	0.031
			0.097	0.072	0.066	0.058	0.046	0.029
$\mathcal{P}_\lambda$	4	-0.195	0.815	0.530	0.378	0.242	0.145	0.046
	8	-1.395	0.696	0.478	0.348	0.220	0.138	0.055
$\mathcal{P}_\pi(0)$	4	-0.117	0.880	0.534	0.377	0.246	0.141	0.049
	8	-0.628	0.867	0.522	0.370	0.259	0.146	0.069
	4,4	-0.172	0.836	0.531	0.376	0.244	0.140	0.048
	8,8	-0.797	0.789	0.520	0.372	0.263	0.148	0.070

<sup>a</sup> $\mathcal{P}_\lambda$ , penalty on canonical eigenvalues;  $\mathcal{P}_\pi(0)$ , penalty on partial correlations, shrinking toward zero with a single value denoting a penalty on genetic correlations only and two values denoting a penalty on both genetic and residual correlations.

<sup>b</sup>Effective sample size of the Beta prior.

<sup>c</sup>Second line gives approximate sampling errors of estimates of canonical eigenvalues.

Experience with applications so far has identified small to moderate effects of penalization on computational requirements compared with an unpenalized analysis, with the bulk of extra computations arising from derivative-free search steps used to check for convergence. The parameterization to the elements of the canonical decomposition, however, tended to increase the number of iterations required even without penalization. Detrimental effects on convergence behavior when parameterizing to eigenvalues of covariance matrices have been reported previously (Pinheiro and Bates 1996).

Convergence rates of iterative maximum-likelihood analyses are dictated by the shape of the likelihood function. Newton–Raphson-type algorithms, including the average information algorithm, involve a quadratic approximation of the likelihood. When this is not the appropriate shape, the algorithm may become “stuck” and fail to locate the maximum. This happens quite frequently for standard (unpenalized) multivariate analyses comprising more than a few traits when estimated covariance matrices have eigenvalues close to zero. For such cases, additional maximization steps using alternative schemes, such as expectation maximization-type algorithms or a derivative-free search, are usually beneficial. For small data sets, we expect the likelihood surface around the maximum to be relatively flat. Adding additional “information” through the assumed prior distribution (*a.k.a.* the penalty) can improve convergence by adding curvature to the surface and creating a more distinct maximum. Conversely, too stringent a penalty may alter the shape of the surface sufficiently so that a quadratic approximation may not be successful. Careful checking of convergence should be an integral part of any multivariate analysis, penalized or not.

## Conclusion

We propose a simple but effective modification of standard multivariate maximum-likelihood analyses to “improve” estimates of genetic parameters: Imposing a penalty on the

likelihood designed to reduce sampling variation will yield estimates that are on average closer to the population values than unpenalized values. There are numerous choices for such penalties. We demonstrate that those derived under the assumption of a Beta distribution for scale-free function of the covariance components to be estimated, namely generalized heritabilities (*a.k.a.* canonical eigenvalues) and genetic correlations, are well suited and tend not to distort estimates of the total, phenotypic variance. In addition, invoking a Beta distribution allows the stringency of penalization to be regulated by a single, intuitive parameter, known as effective sample size of the prior in a Bayesian context. Aiming at moderate rather than optimal improvements in estimates, suitable default values for this parameter can be identified that yield a mild penalty. This allows us to abandon the laborious quest to identify tuning factors suited to particular analyses. Choosing the penalty to be sufficiently mild can all but eliminate the risk of detrimental effects and results in only minor changes in the likelihood, compared to unpenalized analyses. Mildly penalized estimation is recommended for multivariate analyses in quantitative genetics considering more than a few traits to alleviate the inherent effects of sampling variation.

## Acknowledgments

The Animal Genetics and Breeding Unit is a joint venture of the New South Wales Department of Primary Industries and the University of New England. This work was supported by Meat and Livestock Australia grant B.BFG.0050.

## Literature Cited

Anderson, T. W., 1984 *An Introduction to Multivariate Statistical Analysis*, Ed. 2. John Wiley & Sons, New York.

Barnard, J., R. McCulloch, and X. Meng, 2000 Modeling covariance matrices in terms of standard deviations and correlations, with applications to shrinkage. *Stat. Sin.* 10: 1281–1312.

Bickel, P. J., and E. Levina, 2008 Regularized estimation of large covariance matrices. *Ann. Stat.* 36: 199–227.

Bickel, P. J., and B. Li, 2006 Regularization in statistics. *Test* 15: 271–303.

Bouriga, M., and O. Féron, 2013 Estimation of covariance matrices based on hierarchical inverse-Wishart priors. *J. Stat. Plan. Inference* 143: 795–808.

Cheverud, J. M., 1988 A comparison of genetic and phenotypic correlations. *Evolution* 42: 958–968.

Chung, Y., A. Gelman, S. Rabe-Hesketh, J. Liu, and V. Dorie, 2015 Weakly informative prior for point estimation of covariance matrices in hierarchical models. *J. Educ. Behav. Stat.* 40: 136–157.

Daniels, M. J., and R. E. Kass, 2001 Shrinkage estimators for covariance matrices. *Biometrics* 57: 1173–1184.

Daniels, M. J., and M. Pourahmadi, 2009 Modeling covariance matrices via partial autocorrelations. *J. Multivariate Anal.* 100: 2352–2363.

Deng, X., and K.-W. Tsui, 2013 Penalized covariance matrix estimation using a matrix-logarithm transformation. *J. Comput. Graph. Stat.* 22: 494–512.

Fisher, T. J., and X. Sun, 2011 Improved Stein-type shrinkage estimators for the high-dimensional multivariate normal covariance matrix. *Comput. Stat. Data Anal.* 55: 1909–1918.

Gaskins, J. T., M. J. Daniels, and B. H. Marcus, 2014 Sparsity inducing prior distributions for correlation matrices of longitudinal data. *J. Comput. Graph. Stat.* 23: 966–984.

Gelman, A., 2006 Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* 1: 515–533.

Gilmour, A. R., R. Thompson, and B. R. Cullis, 1995 Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51: 1440–1450.

Green, P. J., 1987 Penalized likelihood for general semi-parametric regression models. *Int. Stat. Rev.* 55: 245–259.

Hayes, J. F., and W. G. Hill, 1980 A reparameterisation of a genetic index to locate its sampling properties. *Biometrics* 36: 237–248.

Hayes, J. F., and W. G. Hill, 1981 Modifications of estimates of parameters in the construction of genetic selection indices ('bending'). *Biometrics* 37: 483–493.

Hill, W. G., 2010 Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365: 73–85.

Hsu, C. W., M. S. Sinay, and J. S. J. Hsu, 2012 Bayesian estimation of a covariance matrix with flexible prior specification. *Ann. Inst. Stat. Math.* 64: 319–342.

Huang, A., and M. P. Wand, 2013 Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Anal.* 8: 439–452.

Huang, J. Z., N. Liu, M. Pourahmadi, and L. Liu, 2006 Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika* 93: 85–98.

James, W., and C. Stein, 1961 Estimation with quadratic loss, pp. 361–379 in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, edited by J. Neiman. University of California Press, Berkeley, CA.

Joe, H., 2006 Generating random correlation matrices based on partial correlations. *J. Multivariate Anal.* 97: 2177–2189.

Johnson, N., S. Kotz, and N. Balakrishnan, 1995 *Continuous Univariate Distributions* (Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics Section, Vol. 2, Ed. 2). John Wiley & Sons, New York.

Koots, K. R., J. P. Gibson, and J. W. Wilton, 1994 Analyses of published genetic parameter estimates for beef production traits. 2. Phenotypic and genetic correlations. *Anim. Breed. Abstr.* 62: 825–853.

Lawley, D. N., 1956 Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika* 43: 128–136.

Ledoit, O., and M. Wolf, 2004 A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivariate Anal.* 88: 365–411.

Ledoit, O., and M. Wolf, 2012 Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Stat.* 40: 1024–1060.

Lin, S. P., and M. D. Perlman, 1985 A Monte Carlo comparison of four estimators of a covariance matrix, pp. 411–428 in *Multivariate Analysis*, Vol. 6, edited by P. R. Krishnaiah. North-Holland, Amsterdam.

Meyer, K., 2007 WOMBAT – a tool for mixed model analyses in quantitative genetics by REML. *J. Zhejiang Univ. Sci. B* 8: 815–821.

Meyer, K., 2011 Performance of penalized maximum likelihood in estimation of genetic covariances matrices. *Genet. Sel. Evol.* 43: 39.

Meyer, K., and D. Houle, 2013 Sampling based approximation of confidence intervals for functions of genetic covariance matrices. *Proc. Assoc. Adv. Anim. Breed. Genet.* 20: 523–526.

Meyer, K., and M. Kirkpatrick, 2010 Better estimates of genetic covariance matrices by 'bending' using penalized maximum likelihood. *Genetics* 185: 1097–1110.

- Meyer, K., and S. P. Smith, 1996 Restricted maximum likelihood estimation for animal models using derivatives of the likelihood. *Genet. Sel. Evol.* 28: 23–49.
- Meyer, K., M. Kirkpatrick, and D. Gianola, 2011 Penalized maximum likelihood estimates of genetic covariance matrices with shrinkage towards phenotypic dispersion. *Proc. Assoc. Adv. Anim. Breed. Genet.* 19: 87–90.
- Morita, S., P. F. Thall, and P. Müller, 2008 Determining the effective sample size of a parametric prior. *Biometrics* 64: 595–602.
- Mousseau, T. A., and D. A. Roff, 1987 Natural selection and the heritability of fitness components. *Heredity* 59: 181–197.
- Pinheiro, J. C., and D. M. Bates, 1996 Unconstrained parameterizations for variance-covariance matrices. *Stat. Comput.* 6: 289–296.
- Pourahmadi, M., 2013 *High-Dimensional Covariance Estimation* (Wiley Series in Probability and Statistics). John Wiley & Sons, New York/Hoboken, NJ.
- Powell, M., 2006 The NEWUOA software for unconstrained optimization without derivatives, pp. 255–297 in *Large-Scale Nonlinear Optimization* (Nonconvex Optimization and Its Applications, Vol. 83), edited by G. Pillo and M. Roma. Springer, US.
- Rapisarda, F., D. Brigo, and F. Mercurio, 2007 Parameterizing correlations: a geometric interpretation. *IMA J. Manag. Math.* 18: 55.
- Roff, D. A., 1995 The estimation of genetic correlations from phenotypic correlations - a test of Cheverud's conjecture. *Heredity* 74: 481–490.
- Rothman, A. J., E. Levina, and J. Zhu, 2010 A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* 97: 539–550.
- Schäfer, J., and K. Strimmer, 2005 A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4: 32.
- Stein, C., 1975 Estimation of a covariance matrix, in *Reitz Lecture of the 39th Annual Meeting of the Institute of Mathematical Statistics*. Atlanta, Georgia.
- Thompson, R., S. Brotherstone, and I. M. S. White, 2005 Estimation of quantitative genetic parameters. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360: 1469–1477.
- Waitt, D. E., and D. A. Levin, 1998 Genetic and phenotypic correlations in plants: a botanical test of Cheverud's conjecture. *Heredity* 80: 310–319.
- Warton, D. I., 2008 Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *J. Am. Stat. Assoc.* 103: 340–349.
- Witten, D. M., and R. Tibshirani, 2009 Covariance-regularized regression and classification for high dimensional problems. *J. R. Stat. Soc. B* 71: 615–636.
- Won, J.-H., J. Lim, S.-J. Kim, and B. Rajaratnam, 2013 Condition-number-regularized covariance estimation. *J. R. Stat. Soc. B* 75: 427–450.
- Ye, R. D., and S. G. Wang, 2009 Improved estimation of the covariance matrix under Stein's loss. *Stat. Probab. Lett.* 79: 715–721.
- Zhang, X., W. J. Boscardin, and T. R. Belin, 2006 Sampling correlation matrices in Bayesian models with correlated latent variables. *J. Comput. Graph. Stat.* 15: 880–896.

Communicating editor: L. E. B. Kruuk



## Appendix

### Population Values

Population values for the 13 sets of heritabilities used are summarized in Table A1. The six constellations of genetic ( $r_{G_{ij}}$ ) and residual ( $r_{E_{ij}}$ ) correlations between traits  $i$  and  $j$  ( $i \neq j$ ) were obtained as

- I.  $r_{G_{ij}} = r_{E_{ij}} = 0$ ,
- II.  $r_{G_{ij}} = 0.5$  and  $r_{E_{ij}} = 0.3$ ,
- III.  $r_{G_{ij}} = 0.7^{|i-j|}$  and  $r_{E_{ij}} = 0.5 + 0.05i(-1)^j$ ,
- IV.  $r_{G_{ij}} = -0.7^{|i-j|} + 0.02i$  and  $r_{E_{ij}} = 0.5 + (-0.2)^{|i-j|}$ ,
- V.  $r_{G_{ij}} = r_{E_{ij}} = 0.7$  for  $i, j \in [3, 7]$  and  $r_{G_{ij}} = r_{E_{ij}} = 0.3$  otherwise, and
- VI.  $r_{G_{ij}} = r_{E_{ij}} = 0.6$  for  $|i-j| = 1$ , and  $r_{G_{ij}} = r_{E_{ij}}$  computed from (4) with  $\pi_{ij} = 0.4$  otherwise.

Phenotypic variances were equal to 1 throughout for correlation I and set to 2, 1, 3, 2, 1, 2, 3, 1, and 2 for traits 1–9 otherwise.

### Derivatives of Partial Autocorrelations

Partial autocorrelations [see (3)] can be written as

$$\pi_{ij} = c_{ij} / \sqrt{v_1 v_3}$$

with

$$c_{ij} = r_{ij} - \mathbf{r}'_1 \mathbf{R}_2^{-1} \mathbf{r}_3$$

$$v_x = 1 - \mathbf{r}'_x \mathbf{R}_2^{-1} \mathbf{r}_x \quad \text{for } x = 1, 3.$$

This gives partial derivatives of  $\pi_{ij}$  with respect to parameters  $\theta_k$  and  $\theta_m$ :

$$\frac{\partial \pi_{ij}}{\partial \theta_k} = \frac{\partial c_{ij}}{\partial \theta_k} \frac{\pi_{ij}}{c_{ij}} - \frac{\pi_{ij}}{2} \left( \frac{1}{v_1} \frac{\partial v_1}{\partial \theta_k} + \frac{1}{v_3} \frac{\partial v_3}{\partial \theta_k} \right)$$

$$\begin{aligned} \frac{\partial^2 \pi_{ij}}{\partial \theta_k \partial \theta_m} &= \frac{1}{\sqrt{v_1 v_3}} \left[ \frac{\partial^2 c_{ij}}{\partial \theta_k \partial \theta_m} - \frac{1}{2} \frac{\partial c_{ij}}{\partial \theta_k} \left( \frac{1}{v_1} \frac{\partial v_1}{\partial \theta_m} + \frac{1}{v_3} \frac{\partial v_3}{\partial \theta_m} \right) \right] \\ &+ \frac{\pi_{ij}}{2} \left( \frac{1}{v_1^2} \frac{\partial v_1}{\partial \theta_k} \frac{\partial v_1}{\partial \theta_m} - \frac{1}{v_1} \frac{\partial^2 v_1}{\partial \theta_k \partial \theta_m} + \frac{1}{v_3^2} \frac{\partial v_3}{\partial \theta_k} \frac{\partial v_3}{\partial \theta_m} - \frac{1}{v_3} \frac{\partial^2 v_3}{\partial \theta_k \partial \theta_m} \right) \\ &- \frac{1}{2} \frac{\partial \pi_{ij}}{\partial \theta_m} \left( \frac{1}{v_1} \frac{\partial v_1}{\partial \theta_k} + \frac{1}{v_3} \frac{\partial v_3}{\partial \theta_k} \right). \end{aligned}$$

Derivatives of the components are

$$\frac{\partial c_{ij}}{\partial \theta_k} = \frac{\partial r_{ij}}{\partial \theta_k} + \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_3 - \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_3}{\partial \theta_k} - \frac{\partial \mathbf{r}'_1}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_3$$

$$\begin{aligned} \frac{\partial^2 c_{ij}}{\partial \theta_k \partial \theta_m} &= \frac{\partial^2 r_{ij}}{\partial \theta_k \partial \theta_m} - \frac{\partial^2 \mathbf{r}'_1}{\partial \theta_k \partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_3 - \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial^2 \mathbf{r}_3}{\partial \theta_k \partial \theta_m} + \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial^2 \mathbf{R}_2}{\partial \theta_k \partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_3 - \frac{\partial \mathbf{r}'_1}{\partial \theta_k} \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_3}{\partial \theta_m} - \frac{\partial \mathbf{r}'_1}{\partial \theta_m} \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_3}{\partial \theta_k} \\ &+ \frac{\partial \mathbf{r}'_1}{\partial \theta_k} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_3 + \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_m} \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_3}{\partial \theta_k} + \frac{\partial \mathbf{r}'_1}{\partial \theta_m} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_3 + \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_3}{\partial \theta_m} \\ &- \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_m} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_3 - \mathbf{r}'_1 \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_3 \end{aligned}$$

and

$$\frac{\partial v_x}{\partial \theta_k} = \mathbf{r}'_x \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_x - 2 \mathbf{r}'_x \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_x}{\partial \theta_k}$$

$$\begin{aligned} \frac{\partial^2 v_x}{\partial \theta_k \partial \theta_m} &= \mathbf{r}'_x \mathbf{R}_2^{-1} \frac{\partial^2 \mathbf{R}_2}{\partial \theta_k \partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_x - 2 \frac{\partial^2 \mathbf{r}'_x}{\partial \theta_k \partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_x \\ &+ 2 \frac{\partial \mathbf{r}'_x}{\partial \theta_m} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_x \\ &- 2 \mathbf{r}'_x \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_m} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_k} \mathbf{R}_2^{-1} \mathbf{r}_x \\ &- 2 \frac{\partial \mathbf{r}'_x}{\partial \theta_k} \mathbf{R}_2^{-1} \frac{\partial \mathbf{r}_x}{\partial \theta_m} + 2 \frac{\partial \mathbf{r}'_x}{\partial \theta_k} \mathbf{R}_2^{-1} \frac{\partial \mathbf{R}_2}{\partial \theta_m} \mathbf{R}_2^{-1} \mathbf{r}_x. \end{aligned}$$

Decompose the correlation matrix as  $\mathbf{R} = \mathbf{S}^{-1} \mathbf{\Sigma} \mathbf{S}^{-1}$  with  $\mathbf{S} = \text{Diag}\{s_i\}$  the diagonal matrix of standard deviations for covariance matrix  $\mathbf{\Sigma}$ . This gives the required derivatives of the correlation matrix:

$$\begin{aligned} \frac{\partial \mathbf{R}}{\partial \theta_k} &= \mathbf{S}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \mathbf{S}^{-1} - \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_k} \mathbf{R} - \mathbf{R} \frac{\partial \mathbf{S}}{\partial \theta_k} \mathbf{S}^{-1} \\ \frac{\partial^2 \mathbf{R}}{\partial \theta_k \partial \theta_m} &= \mathbf{S}^{-1} \frac{\partial^2 \mathbf{\Sigma}}{\partial \theta_k \partial \theta_m} \mathbf{S}^{-1} - \mathbf{S}^{-1} \frac{\partial^2 \mathbf{S}}{\partial \theta_k \partial \theta_m} \mathbf{R} - \mathbf{R} \frac{\partial^2 \mathbf{S}}{\partial \theta_k \partial \theta_m} \mathbf{S}^{-1} \\ &- \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_m} \mathbf{S}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \mathbf{S}^{-1} - \mathbf{S}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \theta_k} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_m} \mathbf{S}^{-1} \\ &+ \mathbf{R} \frac{\partial \mathbf{S}}{\partial \theta_k} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_m} \mathbf{S}^{-1} + \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_m} \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_k} \mathbf{R} \\ &- \mathbf{S}^{-1} \frac{\partial \mathbf{S}}{\partial \theta_k} \frac{\partial \mathbf{R}}{\partial \theta_m} - \frac{\partial \mathbf{R}}{\partial \theta_m} \frac{\partial \mathbf{S}}{\partial \theta_k} \mathbf{S}^{-1}. \end{aligned}$$

Finally, assuming derivatives of variances  $\sigma_{ii}$  are available, the required derivatives of standard deviations  $s_i = \sqrt{\sigma_{ii}}$  are

$$\begin{aligned} \frac{\partial s_i}{\partial \theta_k} &= \frac{1}{2s_i} \frac{\partial \sigma_{ii}}{\partial \theta_k} \\ \frac{\partial^2 s_i}{\partial \theta_k \partial \theta_m} &= \frac{1}{2s_i} \left( \frac{\partial^2 \sigma_{ii}}{\partial \theta_k \partial \theta_m} - \frac{1}{2s_i^2} \frac{\partial \sigma_{ii}}{\partial \theta_k} \frac{\partial \sigma_{ii}}{\partial \theta_m} \right). \end{aligned}$$

When estimating elements of  $\mathbf{\Sigma}$  directly, only  $\partial \sigma_{ii} / \partial \sigma_{ii} = 1$  are nonzero. Derivatives of covariance matrices when employing a parameterization to the elements of their Cholesky factors are given by Meyer and Smith (1996).

**Table A1** Population values ( $\times 100$ ) for sets of heritabilities

Set	Trait no.								
	1	2	3	4	5	6	7	8	9
A	40	40	40	40	40	40	40	40	40
B	60	55	50	45	40	35	30	25	20
C	90	60	50	50	30	30	20	20	10
D	75	70	60	50	40	30	20	10	5
E	70	70	70	40	40	40	10	10	10
F	20	20	20	20	20	20	20	20	20
G	35	30	25	20	20	20	15	10	5
H	60	50	10	10	10	10	10	10	10
I	50	50	20	15	15	10	10	5	5
J	80	40	10	10	10	10	10	5	5
K	30	30	25	25	20	15	15	10	10
L	35	30	30	20	20	15	15	15	10
M	10	10	10	30	30	30	50	50	50

# GENETICS

**Supporting Information**

[www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.186114/-/DC1](http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.186114/-/DC1)

## **Simple Penalties on Maximum-Likelihood Estimates of Genetic Parameters to Reduce Sampling Variation**

**Karin Meyer**

# Supplement S1 to “Simple penalties on maximum likelihood estimates of genetic parameters to reduce sampling variation”

## 1 Individual results for sets of population values

Reductions in loss for individual sets of population values showed considerable variation. This is illustrated in a heatmap type plot in Figure S1 for a medium sample size ( $s = 400$  sire families) and an effective sample size of the prior of  $\nu = 8$ . The plot also clearly demonstrates the close relationship between penalties on the improvement in estimates of the genetic covariance matrix and the strong effect the penalty on canonical eigenvalues can have on estimates of the residual covariance matrix.

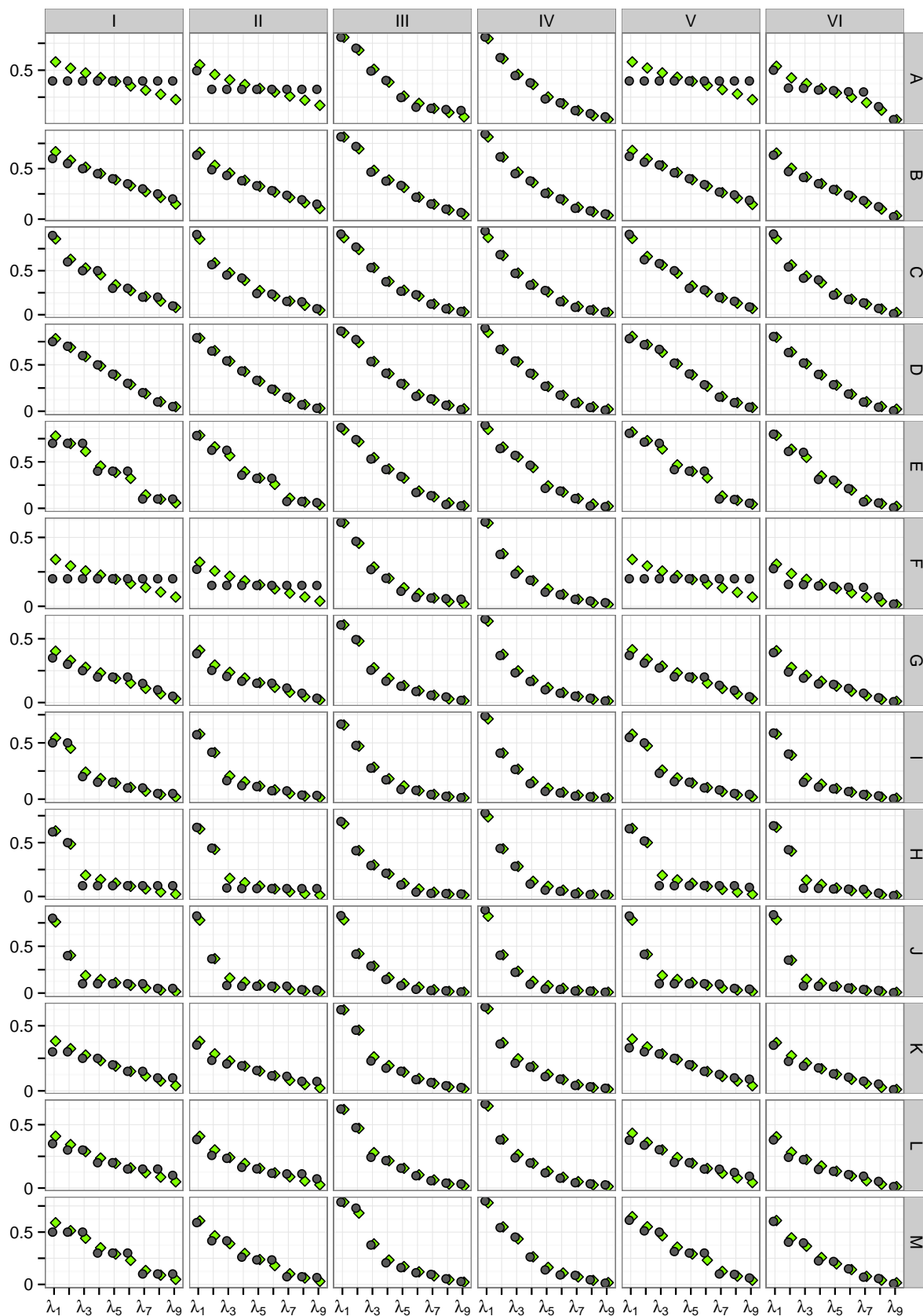
The pattern of PRIAL for  $\hat{\Sigma}_G$  can be directly related to differences in the population values for canonical eigenvalues, summarized in Figure S2. Generally, PRIAL for  $\hat{\Sigma}_G$  tended to be higher for lower levels of heritabilities and thus canonical eigenvalues. Low PRIAL for heritability sets A, B and F combined with correlation structures I, II or V are clearly a reflection of the limited spread in population eigenvalues, which for this sample size does not quite match the range of  $[0, 1]$  assumed in deriving the penalty. As discussed in the manuscript, reducing this interval for penalty  $\mathcal{P}_\lambda$  increased PRIAL for these cases substantially, e.g. from 6 and 7% to 29 and 32% for  $\hat{\Sigma}_G$  and  $\hat{\Sigma}_E$ , respectively, for case A-I and from 19 and 11% to 20 and 26% for case B-V. For the smaller sample ( $s = 100$ , not shown), however, sampling variation for these cases was substantial enough for such ‘range effects’ to be much less evident.

For the remaining cases the spread of eigenvalues appeared sufficient not to clash with the assumption of a distribution over the range  $[0, 1]$ . While differences shown in Figure S2 do not seem to be very pronounced, there are two other outliers, namely cases D-VI and E-VI. Both of these have a wide spread in population eigenvalues, with mean 0.33, the highest value approximately 0.8 and the three smallest values less than 0.1. This constellation appeared to lead to overshrinkage. For more stringent penalties, i.e.  $\nu \geq 12$  these were the only two cases for which PRIAL for  $\hat{\Sigma}_G$  due to penalty  $\mathcal{P}_\lambda$  was negative. It is not clear what particular feature of the population values was responsible for this behaviour – other cases had larger ranges or variances among canonical eigenvalues without being afflicted.

Similar patterns of PRIAL for all penalties (and high correlations shown in the manuscript) suggest that problematic constellations of canonical eigenvalues translate into corresponding difficulties for penalties on correlations. Corresponding modifications of the interval on with PAC are assumed to be distributed may be determined, but have not been investigated so far.

	Genetic													Residual													
VI	47	46	31	4	12	40	30	23	40	29	36	39	26	8	12	59	36	34	9	9	11	14	43	9	9	12	$\mathcal{P}_\lambda$
V	6	10	33	57	58	14	49	56	54	56	41	36	52	7	11	63	34	37	7	9	12	15	41	9	9	13	
IV	50	55	55	33	43	52	43	38	46	38	49	52	42	39	49	40	61	60	12	12	22	26	51	12	13	35	
III	31	52	58	46	55	59	52	47	55	47	56	58	58	36	43	64	60	60	13	13	14	15	40	13	13	36	
II	7	13	45	58	61	31	50	55	57	57	48	46	56	7	11	60	35	33	9	9	12	14	39	9	9	12	
I	6	9	26	54	45	13	47	55	54	57	36	29	41	7	10	63	26	24	7	8	11	13	27	8	8	11	
VI	48	46	31	9	16	47	30	22	44	30	38	42	23	6	5	9	8	10	12	9	7	8	12	9	9	6	$\mathcal{P}_\pi(p)$
V	4	5	29	53	54	22	53	62	61	64	46	41	52	3	2	1	2	2	9	6	5	6	4	7	6	3	
IV	54	59	57	29	39	58	39	31	42	32	51	55	34	18	22	8	15	16	11	10	10	12	21	11	11	17	
III	33	52	59	46	56	68	56	47	57	49	61	64	60	9	11	3	9	8	9	8	7	6	7	9	8	9	
II	6	10	42	57	59	44	58	63	68	67	58	55	57	5	4	5	6	7	13	9	7	8	10	10	10	5	
I	4	4	21	52	41	23	54	63	64	67	45	37	42	4	2	1	2	2	10	7	6	7	5	8	7	3	
VI	48	47	29	3	8	48	24	15	39	24	35	39	20	13	13	26	29	26	17	11	7	7	24	13	12	9	$\mathcal{P}_\pi(0)$
V	5	5	28	54	56	23	55	62	61	64	48	41	54	4	4	6	9	16	9	8	7	8	10	9	8	7	
IV	53	57	57	33	43	60	44	37	47	38	53	57	42	24	32	17	30	29	11	10	11	13	21	10	11	20	
III	34	52	60	46	57	67	56	49	59	51	61	64	60	15	21	11	22	19	9	9	8	8	12	9	9	9	
II	11	13	44	57	61	50	60	64	70	67	61	58	60	11	9	20	21	22	18	11	9	10	15	13	12	9	
I	5	5	19	51	42	23	53	63	64	66	45	37	44	4	2	2	2	2	10	7	6	7	5	8	7	3	
	A	B	C	D	E	F	G	I	H	J	K	L	M	A	B	C	D	E	F	G	I	H	J	K	L	M	

**Figure S1** Percentage reduction in average loss for estimates of genetic and residual covariance matrices for 78 sets of population values for  $s = 400$  sires, applying penalties on canonical eigenvalues ( $\mathcal{P}_\lambda$ ) and partial correlations ( $\mathcal{P}_\pi(0)$ : shrinking towards zero,  $\mathcal{P}_\pi(P)$ : shrinking towards phenotypic values) with an effective prior sample size of  $\nu = 8$



**Figure S2** Population values for canonical eigenvalues ( $\bullet$ ) for 78 sets together with corresponding mean estimates across replicates ( $\blacklozenge$ ) for  $s = 400$  applying a penalty on the canonical eigenvalues ( $\mathcal{P}_\lambda$ ) with an effective prior sample size of  $\nu = 8$



## 2 Detailed results for selected sets of population values

The effect of penalization of estimates of canonical eigenvalues and estimates of genetic correlations is illustrated for two sets of population values, obtained combining heritability set L with the compound symmetric correlation structure, II, (L-II) and combining heritability set H with correlation scenario III (H-III).

### 2.1 Case L-II

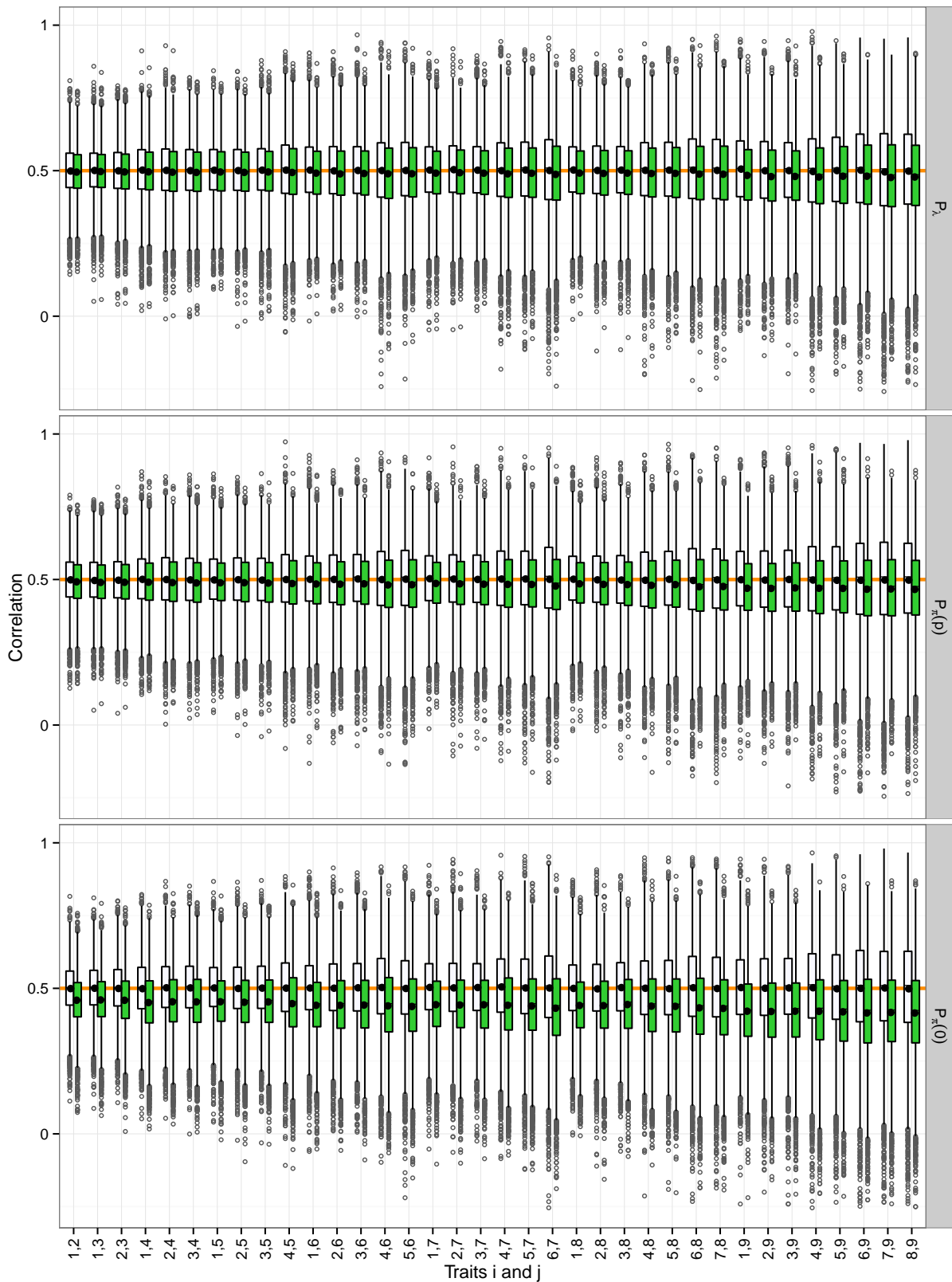
Heritabilities in set L are low to moderate, yielding population canonical eigenvalues ranging from 0.38 to 0.03 with a mean of 0.17. With all genetic correlations in II set to 0.50, PACs at the genetic level depend on the difference between trait numbers only, with values equal to 0.50, 0.33, 0.25, 0.20, 0.17, 0.14, 0.13 and 0.11 for 0 to 8 intervening traits, respectively. Phenotypic correlations for II range from 0.30 to 0.36 with corresponding PAC from 0.36 to 0.07.

Mean estimates of canonical eigenvalues across replicates from unpenalized and penalized analyses with increasingly stringent penalties are summarized in Table S1, together with the respective PRIAL. As expected from sampling theory, unpenalized estimates of eigenvalues are overdispersed with the largest values systematically biased upwards and the smallest values biased downwards. All penalties reduce the overdispersion but in a slightly different fashion. The most notable difference is that the penalty on partial auto-correlations increases the mean eigenvalue slightly, i.e. increases the lowest values more than it reduces the highest ones. For this case, penalties  $\mathcal{P}_\lambda$  and  $\mathcal{P}_\pi(P)$  up  $\nu = 20$  are advantageous. As phenotypic PAC are greater than zero, penalty  $\mathcal{P}_\pi(0)$  is more stringent than  $\mathcal{P}_\pi(P)$ . This is evidenced by larger PRIAL and changes in likelihood at the same PSS. However, a decreasing PRIAL for  $\hat{\Sigma}_E$  for  $\nu > 12$  provides some indication that such values are too stringent – this is supported by the mean estimates of the highest and lowest eigenvalues being below and above, respectively, the corresponding population values.

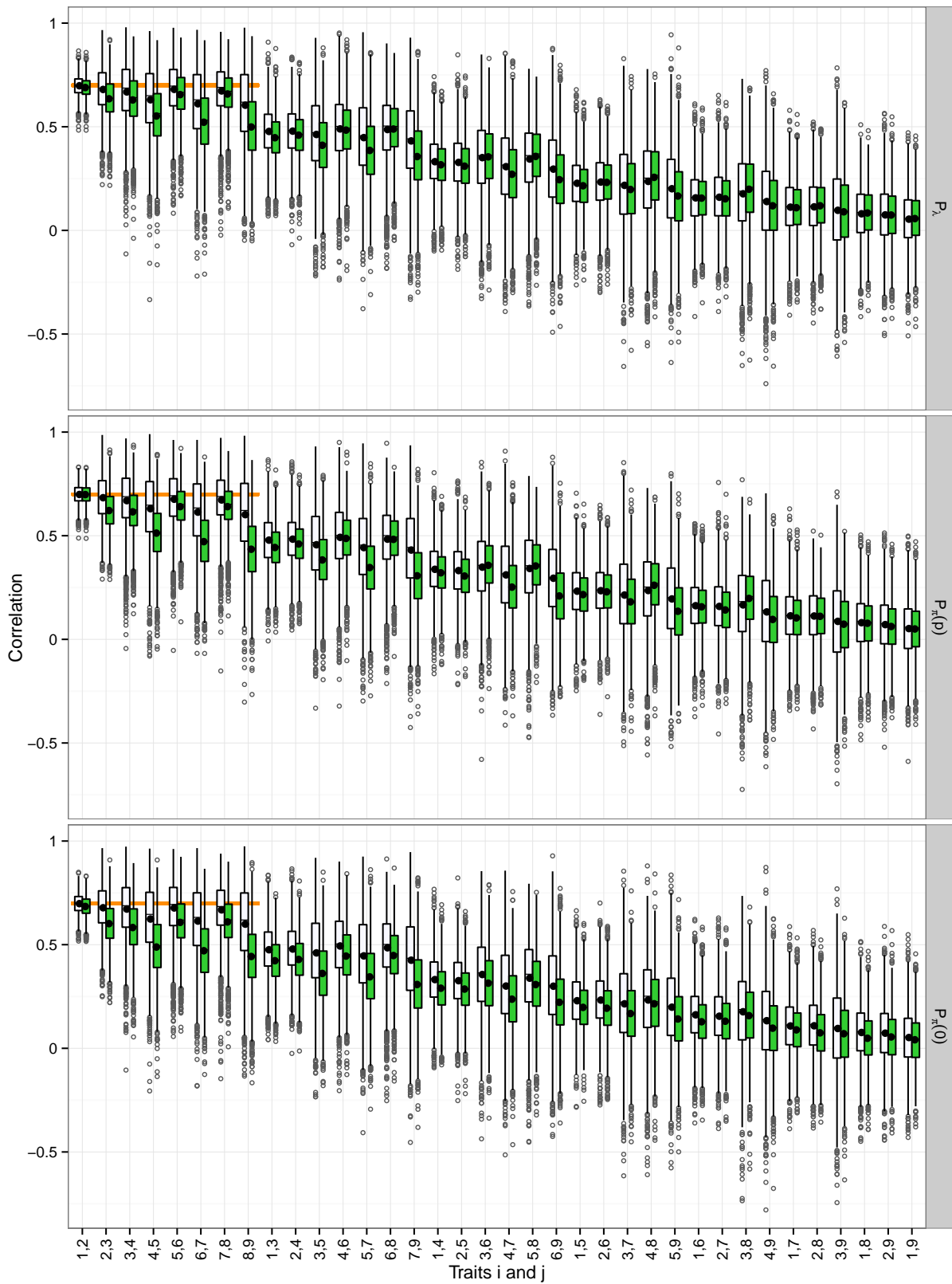
Figure S3 gives the distribution of estimates of genetic correlations, contrasting penalized and unpenalized estimates. Distributions of unpenalized estimates in the three panels are not exactly the same as they represent different simulation runs. Correlations are given in order of the lower triangle of the correlation matrix row-wise. As population heritabilities decrease with trait number, this means that correlations are also shown in increasing ordering of sampling variances. Again, the overall pattern is very similar for all three penalties considered. The effects of penalization are most apparent for correlations for the higher trait numbers. For these, unpenalized estimates are most afflicted by constraints on the parameter space and thus have mean below the population value of 0.5.

### 2.2 Case H-III

Set H is comprised of two high heritabilities, 0.5 and 0.6, with the remainder equal to 0.1. Genetic correlations for III follow an auto-regressive pattern, from 0.7 to  $0.7^8 = 0.06$ . This gives phenotypic correlations for H-III ranging from  $-0.11$  to 0.50, and population canonical eigenvalues from 0.70 to 0.02. The auto-regressive genetic correlation structure



**Figure S3** Distribution of estimates of genetic correlations between trait  $i$  and  $j$  across replicates for population values obtained by combining heritability set  $L$  with correlation scenario II for  $s = 400$ , applying penalties with an effective prior sample size of  $\nu = 8$  (□ unpenalized, ■ penalized estimates; ● mean values )



**Figure S4** Distribution of estimates of genetic correlations between trait  $i$  and  $j$  across replicates for population values obtained by combining heritability set  $H$  with correlation scenario III for  $s = 400$ , applying penalties with an effective prior sample size of  $\nu = 8$  (□ unpenalized, ■ penalized estimates; ● mean values. Horizontal line shows population value for traits  $i$  and  $i + 1$ )

**Table S1** Mean estimates of canonical eigenvalues ( $\times 1000$ ) together with corresponding mean change in unpenalized likelihood for population values obtained combining heritability set L with correlation scenario II for  $s = 400$ .

PSS <sup>a</sup>	Canonical eigenvalues <sup>b</sup>										$\Delta\mathcal{L}^c$	PRIAL <sup>d</sup>	
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\lambda_8$	$\lambda_9$	$\bar{\lambda}$		G	E
	<i>Population values</i>												
	382	257	234	163	152	117	112	112	73	178			
	<i>Unpenalized estimates</i>												
	422	310	250	198	157	120	85	49	13	178	0.000		
	<i>Penalty on canonical eigenvalues</i>												
4	417	308	247	197	156	120	85	51	20	178	-0.070	32	4
6	411	303	245	195	156	120	86	54	23	177	-0.163	40	7
8	409	302	244	195	156	119	87	55	26	177	-0.275	45	10
12	402	298	241	193	155	120	89	59	33	177	-0.478	52	13
16	399	296	240	193	155	121	91	63	37	177	-0.721	57	16
20	393	293	237	192	155	122	93	66	40	177	-0.998	61	19
	<i>Penalty on partial correlations: shrink towards phenotypic</i>												
4	420	308	248	197	157	121	88	55	25	180	-0.175	42	4
6	417	309	248	197	157	122	89	59	32	181	-0.369	51	7
8	412	307	247	197	157	122	90	62	36	181	-0.567	55	10
12	409	307	247	196	158	123	94	67	43	183	-0.952	61	13
16	404	304	244	195	158	126	97	72	49	183	-1.320	65	17
20	400	300	244	196	157	126	99	75	53	183	-1.752	68	19
	<i>Penalty on partial correlations: shrink towards zero</i>												
4	412	309	251	199	157	122	88	55	26	180	-0.238	45	6
6	407	310	250	199	160	124	92	62	33	182	-0.546	53	10
8	399	309	251	200	159	125	94	65	38	182	-0.898	58	13
12	388	306	249	200	163	128	100	73	47	184	-1.863	64	15
16	379	304	250	200	162	129	101	75	52	184	-3.318	66	11
20	370	303	248	199	161	130	103	78	55	183	-4.910	66	7

<sup>a</sup>Effective sample size of prior

<sup>b</sup>In descending order with  $\bar{\lambda}$  denoting their mean

<sup>c</sup>Change in log likelihood compared to unpenalized analysis

<sup>d</sup>Percentage reduction in average loss for genetic (G) and residual (E) covariance matrix

gives genetic PAC for H-III that are high for 'adjacent' traits (0.7) and zero otherwise. Corresponding phenotypic values range from 0.50 (traits 1 and 2) to 0.11, creating differences between genetic and phenotypic PAC up to 0.81.

Mean estimates of canonical eigenvalues for different penalties for this set of population values are given in Table S2 and the distribution of estimates of genetic correlations is shown in Figure S4. In this case, correlations are arranged according to their population values, with groups equal to the elements of successive subdiagonal of the lower triangle of the correlation matrix. While the general pattern of results is similar to that observed for population values L-II, the spread of canonical eigenvalues is larger with more eigenvalues close to zero and the effects of sampling variation are thus more pronounced.

**Table S2** Mean estimates of canonical eigenvalues ( $\times 1000$ ) together with corresponding mean change in unpenalized likelihood for population values obtained combining heritability set H with correlation scenario III for  $s = 400$ .

PSS <sup>a</sup>	Canonical eigenvalues <sup>b</sup>										$\Delta\mathcal{L}^c$	PRIAL <sup>d</sup>	
	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	$\lambda_5$	$\lambda_6$	$\lambda_7$	$\lambda_8$	$\lambda_9$	$\bar{\lambda}$		G	E
	<i>Population values</i>												
	696	425	287	215	108	41	29	26	23	206			
	<i>Unpenalized estimates</i>												
	715	445	305	214	121	67	29	5	1	211			
	<i>Penalty on canonical eigenvalues</i>												
4	699	436	302	209	119	69	34	12	5	209	-0.326	40	10
6	688	433	296	211	120	70	37	17	8	209	-0.712	49	13
8	678	431	296	209	118	71	41	21	12	208	-1.115	54	15
12	664	422	290	207	124	75	46	26	17	208	-1.974	58	17
16	648	415	289	207	123	76	49	30	21	206	-2.888	59	15
20	632	412	289	206	125	81	54	36	25	206	-3.698	57	12
	<i>Penalty on partial correlations: shrink towards phenotypic</i>												
4	719	445	302	211	122	69	38	18	10	215	-0.739	52	3
6	706	438	300	209	123	74	44	25	16	215	-1.447	57	4
8	712	436	302	208	123	76	50	31	21	218	-2.026	57	6
12	706	429	301	203	123	81	56	38	28	218	-3.260	57	8
16	702	427	297	205	123	84	61	44	34	220	-4.361	54	11
20	697	425	294	202	126	88	66	50	40	221	-5.343	51	11
	<i>Penalty on partial correlations: shrink towards zero</i>												
4	710	438	303	213	120	69	38	18	10	213	-0.683	51	3
6	709	444	304	207	121	73	44	25	15	216	-1.340	57	5
8	701	440	306	205	124	75	47	29	20	216	-1.997	59	7
12	697	438	303	199	124	79	54	37	27	218	-3.331	58	10
16	689	440	299	194	125	81	58	41	31	218	-4.863	57	10
20	679	441	303	195	125	84	60	45	35	219	-6.267	55	10

<sup>a</sup>Effective sample size of prior

<sup>b</sup>In descending order with  $\bar{\lambda}$  denoting their mean

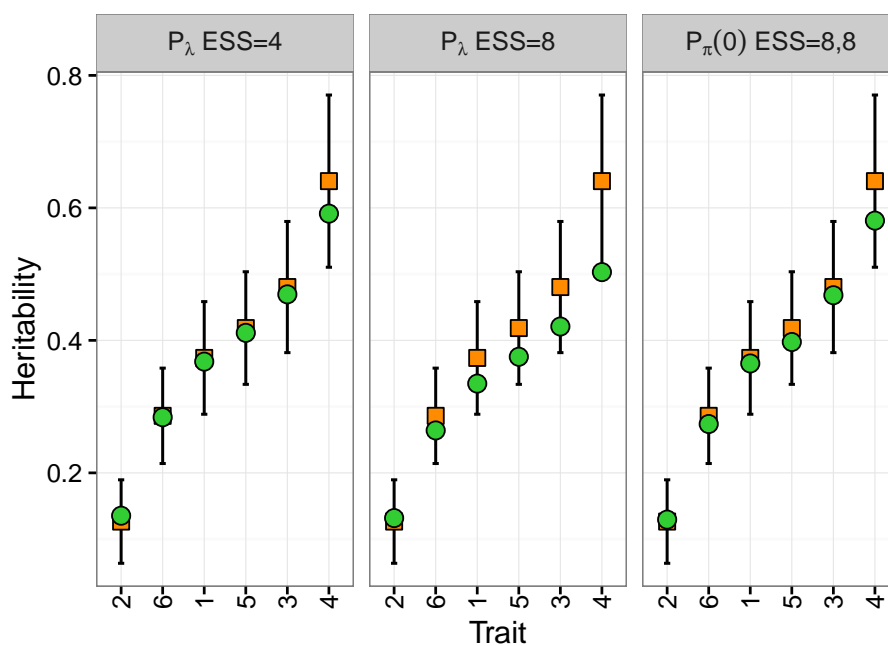
<sup>c</sup>Change in log likelihood compared to unpenalized analysis

<sup>d</sup>Percentage reduction in average loss for genetic (G) and residual (E) covariance matrix

In particular, bias in unpenalized estimates of genetic correlations between traits  $i$  and  $i + 1$  due to constraints on the parameter space is notable for the traits with low heritability ( $i \geq 3$ ). This bias is larger if the corresponding residual correlation is negative. This results in a substantial difference between genetic and phenotypic values so that penalization increases bias in estimates of those genetic correlations quite markedly.

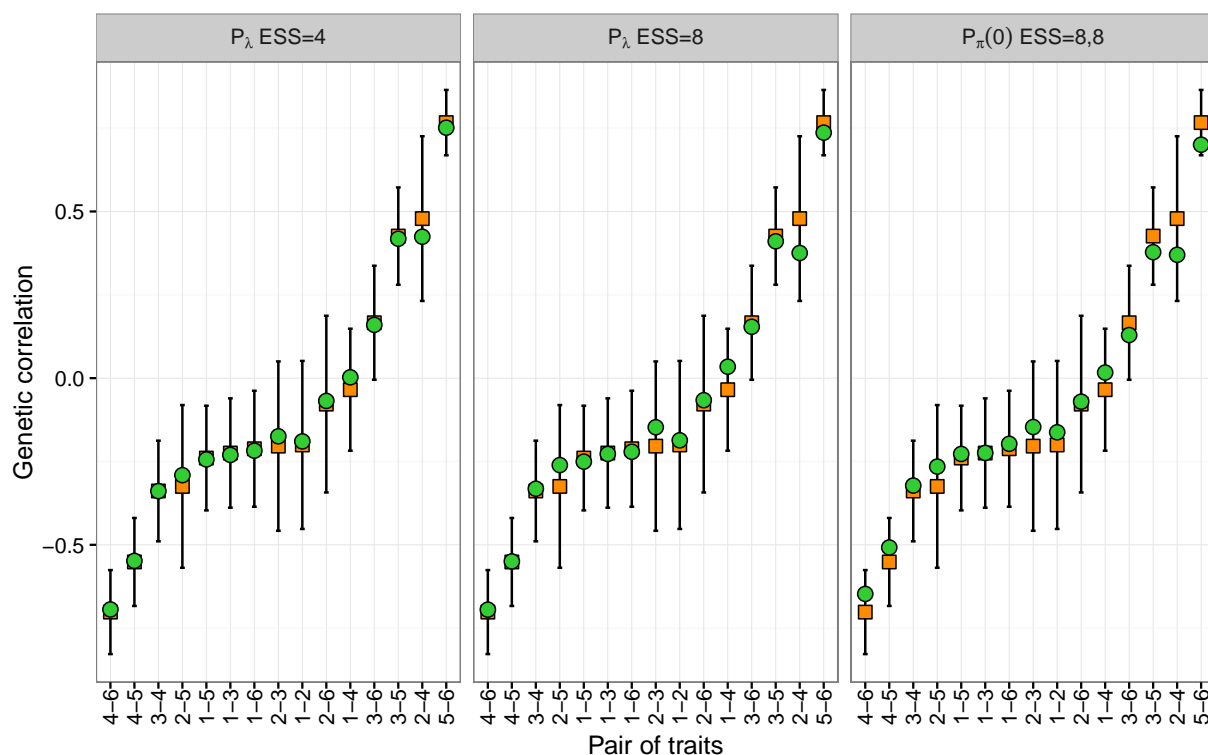
### 3 Detailed results for applied example

The effect of penalized estimation for the applied example on individual genetic parameters is shown in Figure S5, Figure S6 and Figure S7 for heritabilities, genetic and residual correlations, respectively, showing values in ascending order of unpenalized estimates. Figures demonstrate that the effect of penalization was generally very small. Estimates that changed most, by approximately one standard error, were the heritability for trait 4, the trait with the least number of observation, and the residual correlation between traits 1 and 4, for a penalty on canonical eigenvalues with PSS of 8.

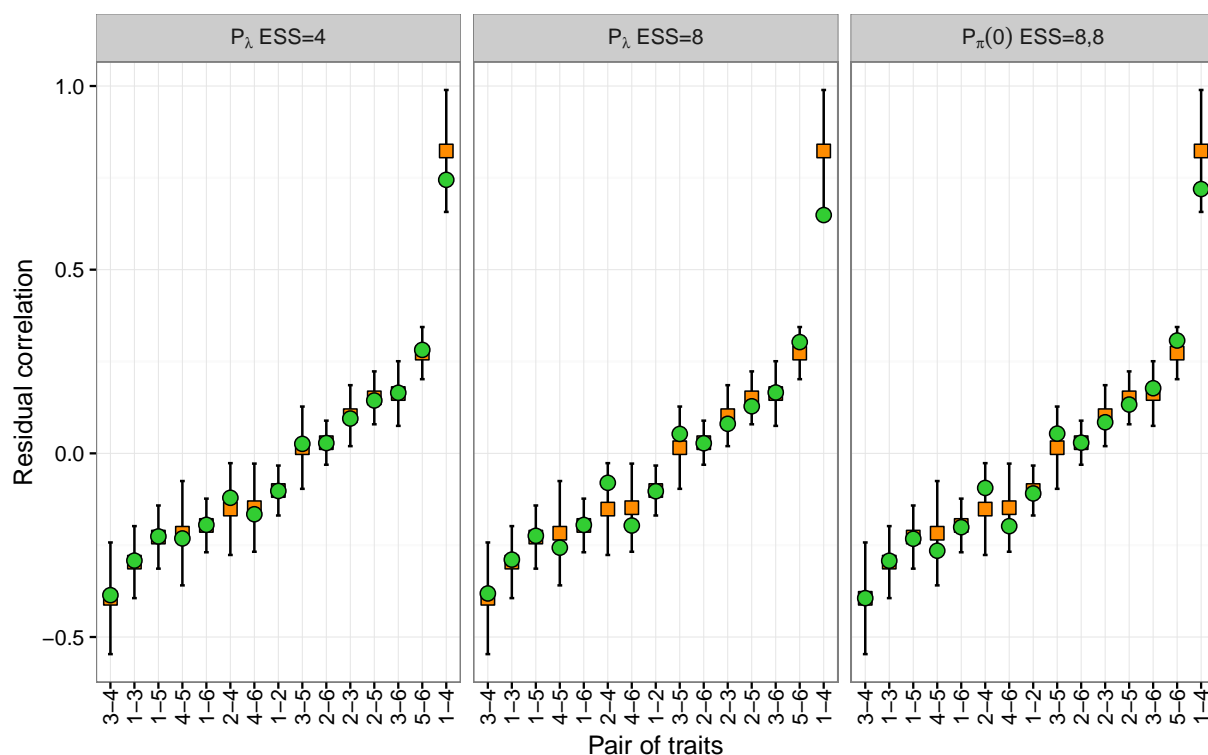


**Figure S5** Effect of penalization on estimates of individual heritabilities (■ Unpenalized ● Penalized; vertical bars shown the range of plus/minus one standard deviation for unpenalized estimates)





**Figure S6** Effect of penalization on estimates of individual genetic correlations (■ Unpenalized ● Penalized; vertical bars shown the range of plus/minus one standard deviation for unpenalized estimates)



**Figure S7** Effect of penalization on estimates of individual residual correlations (■ Unpenalized ● Penalized; vertical bars shown the range of plus/minus one standard deviation for unpenalized estimates)

**Table S3** Characteristics of convergence behaviour and changes in likelihood for a single, simulated data set for  $s = 400$  sires with population parameters obtained combining heritability set H with correlation scenario III.

PSS <sup>a</sup>		Unpenalized		$\mathcal{P}_\lambda$	$\mathcal{P}_\pi(0)$	$\mathcal{P}_\pi(P)$
		Cholesky <sup>b</sup>	Canonical <sup>c</sup>	Canonical	Cholesky	Cholesky
8	Likelihood <sup>d</sup>	0	-0.002	-1.397	-1.742	-1.578
	AI iterates <sup>e</sup>	10	25	30	10	8
	DF iterates <sup>f</sup>	0.000	0.000	0.001	0.058	0.060
20	Likelihood			-8.252	-5.979	-4.294
	AI iterates			25	11	9
	DF iterates			0.000	0.079	0.109

<sup>a</sup>Effective sample size  $\nu$

<sup>b</sup>Standard parameterisation to elements of the Cholesky factor of covariance matrices to be estimated

<sup>c</sup>Parameterisation to elements of the canonical transformation

<sup>d</sup>Unpenalized log likelihood, expressed as deviation from standard, unpenalized analysis

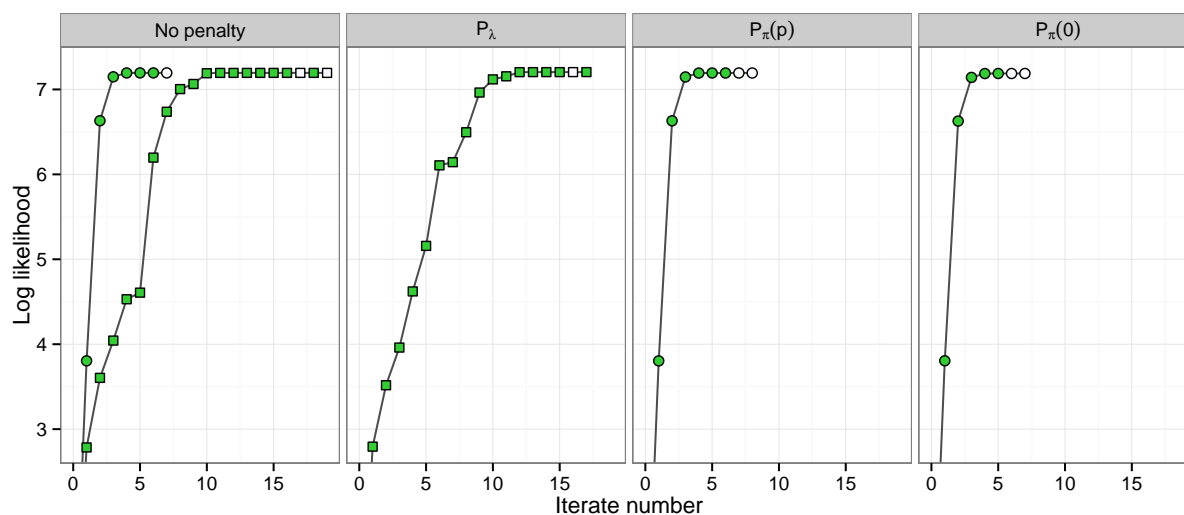
<sup>e</sup>Number of average information algorithm REML iterates carried out

<sup>f</sup>Change in penalized log likelihood in subsequent derivative-free maximization steps, expressed as deviation from value for last AI step

## 4 Effect of penalization on convergence of REML analyses

We illustrate the effects of penalization on the convergence behaviour of REML analyses for a data set simulated for population values H-III and  $s = 400$  sires and effective sample sizes of  $\nu = 8$  and  $\nu = 20$ . Analyses were carried out using our mixed model package WOMBAT, using an average information algorithm to locate the maximum of the likelihood, followed by derivative-free search steps (and possibly additional average information and derivative-free steps) to ensure convergence. Data and parameter files are available as worked Example 19 for the package.

Figure S8 shows the convergence behaviour for mild penalization ( $\nu = 8$ ) and additional characteristics are summarized in Table S3. For a simple model with means as the only fixed effects and all records recorded on all animals, convergence of the standard, unpenalized analysis is rapid in spite of the substantial number of parameters to be estimated (45 genetic covariances and 45 residual covariances). This uses a parameterisation to the elements of the Cholesky factors of the two covariance matrices to be estimated. Performing the same analysis parameterising to estimate the canonical eigenvalues and elements of the canonical transformation instead increases the number of iterates required. The effect of parameterisation carries through to penalized analyses. While estimation on the canonical simplifies implementation, it is relatively slow to converge. Convergence appeared to be to the maximum of the penalized likelihood, with derivative-free search steps not increasing it further, even for more stringent penalization for  $\nu = 20$ . In contrast, penalties on correlations are implemented using the standard parameterisation to elements of the Cholesky factor. Thus penalized estimation has little impact on the number of average information algorithm iterates required. However, for these cases subsequent derivative-free search steps yield additional, small increases in the likelihood, especially for the more stringent penalisation for  $\nu = 20$ .



**Figure S8** Effect of penalization and parameterisation on convergence of REML analyses for a single replicate of simulated data obtained by combining heritability set H with correlation scenario III for  $s = 400$ , applying penalties with an effective prior sample size of  $\nu = 8$  (Parameterisation: ● Elements of Cholesky factors of covariance matrices, ■ Elements of canonical decomposition; filled symbols: average information, open symbols: derivative-free maximization steps; Likelihood values are given as deviation from value starting values and scaled by dividing by 1000)