Journal of
Animal Breeding and Genetics

J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Bayesian conjugate analysis using a generalized inverted Wishart distribution accounts for differential uncertainty among the genetic parameters – an application to the maternal animal model

S. Munilla[1] & R.J.C. Cantet[1,2]

1 Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina
2 Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

Keywords
Elicitation methods; gibbs sampler; prior distributions; variance components estimation.

Correspondence
Sebastián Munilla, Av. San Martín 4453 (C1417DSE), Buenos Aires, Argentina.
Tel: 54 11 4524 8000, ext. 8192;
Fax: 54 11 4524 8735; E-mail: munilla@agro.uba.ar

Received: 21 February 2011;
accepted: 22 July 2011

Summary

Consider the estimation of genetic (co)variance components from a maternal animal model (MAM) using a conjugated Bayesian approach. Usually, more uncertainty is expected a priori on the value of the maternal additive variance than on the value of the direct additive variance. However, it is not possible to model such differential uncertainty when assuming an inverted Wishart (IW) distribution for the genetic covariance matrix. Instead, consider the use of a generalized inverted Wishart (GIW) distribution. The GIW is essentially an extension of the IW distribution with a larger set of distinct parameters. In this study, the GIW distribution in its full generality is introduced and theoretical results regarding its use as the prior distribution for the genetic covariance matrix of the MAM are derived. In particular, we prove that the conditional conjugacy property holds so that parameter estimation can be accomplished via the Gibbs sampler. A sampling algorithm is also sketched. Furthermore, we describe how to specify the hyperparameters to account for differential prior opinion on the (co)variance components. A recursive strategy to elicit these parameters is then presented and tested using field records and simulated data. The procedure returned accurate estimates and reduced standard errors when compared with non-informative prior settings while improving the convergence rates. In general, faster convergence was always observed when a stronger weight was placed on the prior distributions. However, analyses based on the IW distribution have also produced biased estimates when the prior means were set to over-dispersed values.

Introduction

This article deals with the use of the generalized inverted Wishart (GIW) distribution for tackling the problem of estimating genetic covariance matrices under a conjugated Bayesian approach. The GIW distribution was originally introduced by Brown *et al.* (1994) in the context of risk assessment of air pollution (cf. Le & Zidek 2006), and it is essentially an extension of the inverted Wishart (IW) distribution with a larger set of distinct parameters, a feature that confers upon the density a great flexibility. In particular, the GIW distribution arises as a natural specification for the prior covariance structure of

  doi:10.1111/j.1439-0388.2011.00953.x

multivariate Gaussian data with a monotone pattern of missing values (Garthwaite & Al-Awadhi 2001).

In connection with this, it is argued that the distribution could also be used to specify a more flexible prior genetic covariance structure in the context of the statistical models used by animal breeders, in particular when differential information is available for the estimation of distinct scalar components. Take, for instance, the maternal animal model (MAM), where usually more uncertainty (less information) is expected on the estimated maternal additive variance or maternal heritability than on the direct additive variance or direct heritability. In such case, the analyst would reasonably tend to favour a different prior specification to represent this differential uncertainty.

The objective of this research is threefold. First, the GIW distribution will be formally introduced in its full generality. Next, we present theoretical results regarding the use of the GIW as the prior distribution for the genetic covariance matrix of the MAM under a hierarchical Bayesian analysis. Finally, we describe a Bayesian updating approach to elicit prior parameters, taking advantage of the properties of the GIW distribution, and further estimate genetic parameters using both field weaning weight records and simulated data. Results are later compared with more standard prior specifications, focusing on the accuracy of the estimates, their standard errors and the convergence behaviour of the Markov chains.

Methods

The generalized inverted Wishart distribution

Let $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ be a sample of g-dimensional data vectors and define $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$. Consider now a Gaussian matrix-variate distribution for the random matrix $\boldsymbol{Y}$ $(n \times g)$ thus formed, such that

$$\mathrm{vec}(\boldsymbol{Y})|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{A}), \tag{1}$$

where $\mathrm{vec}(\cdot)$ denotes the vec operator (Searle 1982, ch. 12.9), $\boldsymbol{A}$ is a $(n \times n)$ known symmetric matrix and $\boldsymbol{\Sigma}$ is a random $(g \times g)$ covariance matrix subject to a hierarchical structure. Under this setting, the latter is usually assumed to follow an IW distribution a priori, i.e. $\boldsymbol{\Sigma} \sim IW(\delta, \boldsymbol{\Psi})$. In this respect, note that while the symmetric positive definite matrix $\boldsymbol{\Psi}$ provides a full set of complementary hyperparameters to assess the prior expectation, the uncertainty attached to it is governed by a single scalar parameter, δ (Brown 2002). Instead, a more flexible approach can be obtained through the Bartlett decomposition (Bartlett 1933) of $\boldsymbol{\Sigma}$.

Consider first a partition of the covariance matrix in 2×2 blocks,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}. \tag{2}$$

The Bartlett decomposition of $\boldsymbol{\Sigma}$ is such that $\boldsymbol{\Sigma} = \boldsymbol{T}\Delta\boldsymbol{T}'$, with

$$\Delta = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \end{bmatrix} \quad \text{and} \quad \boldsymbol{T} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} & \boldsymbol{I} \end{bmatrix}. \tag{3}$$

Denote $\boldsymbol{\tau} \equiv \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$ and $\boldsymbol{\Gamma} \equiv \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$. Then, the decomposition could be regarded as a one-to-one transformation of the covariance matrix $\boldsymbol{\Sigma} \rightarrow (\boldsymbol{\Sigma}_{11}, \boldsymbol{\tau}, \boldsymbol{\Gamma})$, such that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\boldsymbol{\tau}' \\ \boldsymbol{\tau}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Gamma} + \boldsymbol{\tau}\boldsymbol{\Sigma}_{11}\boldsymbol{\tau}' \end{bmatrix}. \tag{4}$$

In the more general case of multiple blocks, the decomposition could be applied recursively. However, before proceeding further, we need to establish some notation. Throughout the article, we will adopt the one by Le *et al.* (1999).

Let $\boldsymbol{\Sigma}$ be arbitrarily partitioned as having $k \times k$ blocks,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \cdots & \boldsymbol{\Sigma}_{1,k} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{k,1} & \cdots & \boldsymbol{\Sigma}_{k,k} \end{bmatrix}, \tag{5}$$

with $\boldsymbol{\Sigma}_{i,l}$ having dimensions $g_i \times g_l$, such that $g_1 + \cdots + g_k = g$. Now, denote the leading principal submatrix up to the jth block as $\boldsymbol{\Sigma}^{[1,\ldots,j]}$. That is,

$$\boldsymbol{\Sigma}^{[1,\ldots,j]} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \cdots & \boldsymbol{\Sigma}_{1,j} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{j,1} & \cdots & \boldsymbol{\Sigma}_{j,j} \end{bmatrix}. \tag{6}$$

Further, let $\boldsymbol{\Sigma}^{[(j+1)j]} = (\boldsymbol{\Sigma}_{(j+1),1}, \ldots, \boldsymbol{\Sigma}_{(j+1),j})$ and $\boldsymbol{\Sigma}^{[j(j+1)]} = (\boldsymbol{\Sigma}^{[(j+1)j]})'$, the latter based on the symmetry of $\boldsymbol{\Sigma}$. Then, the Bartlett decomposition could be applied recursively, as for $j = k - 1, \ldots, 1$

$$\boldsymbol{\Sigma}^{[1,\ldots,j+1]} = \begin{bmatrix} \boldsymbol{\Sigma}^{[1,\ldots,j]} & \boldsymbol{\Sigma}^{[1,\ldots,j]}\boldsymbol{\tau}_j' \\ \boldsymbol{\tau}_j\boldsymbol{\Sigma}^{[1,\ldots,j]} & \boldsymbol{\Gamma}_j + \boldsymbol{\tau}_j\boldsymbol{\Sigma}^{[1,\ldots,j]}\boldsymbol{\tau}_j' \end{bmatrix}, \tag{7}$$

with $\boldsymbol{\tau}_j = \boldsymbol{\Sigma}^{[(j+1)j]}\left(\boldsymbol{\Sigma}^{[1,\ldots,j]}\right)^{-1}$ and $\boldsymbol{\Gamma}_j = \boldsymbol{\Sigma}_{(j+1),(j+1)} - \boldsymbol{\Sigma}^{[(j+1)j]}\left(\boldsymbol{\Sigma}^{[1,\ldots,j]}\right)^{-1}\boldsymbol{\Sigma}^{[j(j+1)]}$.

Consider now a conformable partition of the data matrix $\boldsymbol{Y}$ into k blocks,

$$\boldsymbol{Y} = \left(\boldsymbol{Y}^{[1]}, \ldots, \boldsymbol{Y}^{[k]}\right) \tag{8}$$

with $\boldsymbol{Y}^{[i]} = \left(\boldsymbol{y}_1^{[i]}, \ldots, \boldsymbol{y}_n^{[i]}\right)'$ for the ith block. The blocks have dimension $n \times g_l$, such that $g_1 + \cdots + g_k = g$. The notation stresses the fact that there is not necessarily a one-to-one correspondence between a data block and each coordinate of the vectors $\boldsymbol{y}$'s, although for convenience we shall assume so in future developments, so that $g_l = 1$ for all l and thus $\boldsymbol{Y}^{[i]}$ could be regarded as a vector.

Now, using properties of the normal distribution (cf. Bauwens *et al.* 1999, Appendix A.2.3) and the notation from the multiple-block Bartlett decomposition aforementioned, we can express the joint density of $\boldsymbol{Y}$ as the product of the following sequence of conditional distributions (Brown 2002)

$$\begin{aligned} \boldsymbol{Y}^{[1]} &\sim N(\boldsymbol{0}, \boldsymbol{A} \otimes \boldsymbol{\Sigma}_{1,1}) \\ \boldsymbol{Y}^{[2]} | \boldsymbol{Y}^{[1]} &\sim N(\boldsymbol{Y}^{[1]}\boldsymbol{\tau}_1, \boldsymbol{A} \otimes \boldsymbol{\Gamma}_1) \\ &\vdots \\ \boldsymbol{Y}^{[j+1]} | \boldsymbol{Y}^{[1]}, \ldots, \boldsymbol{Y}^{[j]} &\sim N(\boldsymbol{Y}^{[1,\ldots,j]}\boldsymbol{\tau}_j, \boldsymbol{A} \otimes \boldsymbol{\Gamma}_j), \end{aligned} \tag{9}$$

for $j = 2, \ldots, k-1$, with $\boldsymbol{Y}^{[1,\ldots,j]} = \left(\boldsymbol{Y}^{[1]}, \ldots, \boldsymbol{Y}^{[j]}\right)$.

We have now arrived at the main objective of this section: the definition of the GIW distribution. Notice that Equation (9) gives an insight into how to set a more parameterized prior distribution for the covariance matrix $\boldsymbol{\Sigma}$. Based on the mutual independence between $\boldsymbol{\Sigma}_{1,1}$ and pairs $\left(\boldsymbol{\tau}_j, \boldsymbol{\Gamma}_j\right)$, $j = 1, \ldots, k-1$, a property that rests on the Bartlett decomposition, assume that a priori

$$\begin{aligned} \boldsymbol{\Sigma}_{1,1} &\sim IW(\delta_0, \boldsymbol{Q}_0) \\ \boldsymbol{\tau}_j | \boldsymbol{\Gamma}_j &\sim N\left(\boldsymbol{\tau}_{0j}, \boldsymbol{H}_j \otimes \boldsymbol{\Gamma}_j\right) \\ \boldsymbol{\Gamma}_j &\sim IW\left(\delta_j + g^{[1,\ldots,j]}, \boldsymbol{Q}_j\right), \end{aligned} \tag{10}$$

with $g^{[1,\ldots,j]} = g_1 + \cdots + g_j$. In Equation (10), $\{\delta_0, \boldsymbol{Q}_0, \delta_j, \boldsymbol{\tau}_{0j}, \boldsymbol{Q}_j, \boldsymbol{H}_j,\ j = 1, \ldots, k-1\}$ embodies a set $\mathscr{H}$ of hyperparameters. Then, it is said that the covariance matrix $\boldsymbol{\Sigma}$ follows a $GIW(\mathscr{H})$ distribution (Brown 2002).

The GIW distribution is essentially characterized by the larger set of parameters it involves, a feature that offers great flexibility while specifying prior knowledge or expert opinion regarding the covariance structure of the data. Furthermore, it has the advantage of computation simplicity as the Bartlett parameters follow distributions easy to sample from, namely Gaussian and lower-order IW densities. Finally, notice the GIW is equivalently defined in a recursive fashion, as the principal leading submatrix $\boldsymbol{\Sigma}^{[1,\ldots,j]}$ of $\boldsymbol{\Sigma}$ can be regarded as being distributed

$$\boldsymbol{\Sigma}^{[1,\ldots,j]} \sim GIW(\delta_i, \boldsymbol{Q}_i, \boldsymbol{\tau}_{0i}, \boldsymbol{H}_i,\ i = 1, \ldots, j-1) \tag{11}$$

successively for $j = k-1, \ldots, 2$, while setting $\boldsymbol{\Sigma}^{[1,1]} \equiv \boldsymbol{\Sigma}_{1,1} \sim IW(\delta_0, \boldsymbol{Q}_0)$ for $j = 1$.

Prior specification using the GIW: theoretical results

Our goal now is to develop some theory on the use of the GIW distribution in the context of a hierarchical Bayesian analysis of Gaussian performance data. In particular, consider the problem of estimating the genetic (co)variance components for the additive genetic model with maternal effects through a Gibbs sampler (cf. Sorensen & Gianola 2002, ch. 13.3). Here, it is standard to assume a priori an IW distribution for the genetic covariance matrix. Yet, we will show that setting a GIW distribution instead broadens considerably the range of possible prior specification, while keeping the main advantage of the approach: conditional conjugacy. We will first introduce the model and some basic notation.

The maternal animal model

Maternal animal models (MAMs) are mixed linear models used to fit records on maternally influenced traits. The additive covariance structure in these models is based on the theory of the covariance among relatives of Willham (1963), and its formulation within the mixed linear models theory is indebted to Quaas & Pollak (1980). On using subscripts 'o' and 'm' to differentiate between the direct and the maternal effects, respectively, we can express the model equation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}_{\mathrm{o}}\boldsymbol{a}_{\mathrm{o}} + \boldsymbol{Z}_{\mathrm{m}}\boldsymbol{a}_{\mathrm{m}} + \boldsymbol{Z}_{\mathrm{p}}\boldsymbol{e}_{\mathrm{p}} + \boldsymbol{e}_{\mathrm{o}} \tag{12}$$

where $\boldsymbol{y}$ $(n \times 1)$ is a data vector and $\boldsymbol{X}$ $(n \times p)$ is the incidence matrix for the fixed effects vector $\boldsymbol{b}$ $(p \times 1)$. Additionally, $\boldsymbol{a}_{\mathrm{o}}$ and $\boldsymbol{a}_{\mathrm{m}}$ are $(q \times 1)$ random vectors with entries corresponding to the direct and maternal breeding values, respectively, and $\boldsymbol{e}_{\mathrm{p}}$ $(d \times 1)$ is a random vector accounting for maternal permanent environmental effects. Accordingly, $\boldsymbol{Z}_{\mathrm{o}}$, $\boldsymbol{Z}_{\mathrm{m}}$ and $\boldsymbol{Z}_{\mathrm{p}}$ are the corresponding incidence matrices. Finally, $\boldsymbol{e}_{\mathrm{o}}$ $(n \times 1)$ is a random vector of errors. To simplify the notation, let $\boldsymbol{a}' = \left(\boldsymbol{a}_{\mathrm{o}}', \boldsymbol{a}_{\mathrm{m}}'\right)$.

All random effects are defined as deviations from their mean values, and thus, their expectation is 0.

The model is then completed with the following covariance specification

$$\mathrm{Cov}\begin{bmatrix}\boldsymbol{a}\\ \boldsymbol{e}_{\mathrm{p}}\\ \boldsymbol{e}_{\mathrm{o}}\end{bmatrix}=\begin{bmatrix}\Sigma\otimes\boldsymbol{A} & 0 & 0\\ 0 & \boldsymbol{I}_d\sigma^2_{e_{\mathrm{m}}} & 0\\ 0 & 0 & \boldsymbol{I}_n\sigma^2_{e_{\mathrm{o}}}\end{bmatrix}, \qquad (13)$$

where $\boldsymbol{A}$ is the $(q\times q)$ numerator relationship matrix and $\boldsymbol{\Sigma}$ is a (2×2) matrix containing scalar genetic (co)variance components, i.e.

$$\boldsymbol{\Sigma}=\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\equiv\begin{bmatrix}\sigma^2_{a_{\mathrm{o}}} & \sigma_{a_{\mathrm{o}}a_{\mathrm{m}}}\\ \sigma_{a_{\mathrm{o}}a_{\mathrm{m}}} & \sigma^2_{a_{\mathrm{m}}}\end{bmatrix}. \qquad (14)$$

The estimation of the latter parameters under a hierarchical Bayesian approach requires the specification of a prior distribution for matrix $\boldsymbol{\Sigma}$. In this respect, usually an IW distribution is set, as it turns out that the full conditional distribution belongs to the same family, a property known as conditional conjugacy (Daniels & Pourahmadi 2002). In fact, the conditional conjugacy property is one of the more attractive features of the IW distribution, as it enables the use of well-known sampling algorithms for estimation purposes, such as the Gibbs sampler (Gelfand & Smith 1990). A more detailed description of the Bayesian analysis for maternal influenced traits can be found in Cantet *et al.* (1992), whereas details in the estimation of the (co)variance components through a Gibbs sampler are in the study by Jensen *et al.* (1994).

Partition of the vector of breeding values

Let the prior distribution of the vector of breeding values in the MAM [Equation (12)] be

$$\boldsymbol{a}=\begin{bmatrix}\boldsymbol{a}_{\mathrm{o}}\\ \boldsymbol{a}_{\mathrm{m}}\end{bmatrix}\sim N(\boldsymbol{0},\boldsymbol{\Sigma}\otimes\boldsymbol{A}). \qquad (15)$$

Based on the results presented in the previous section, consider expressing this joint density as the product of the following distributions

$$\begin{aligned}\boldsymbol{a}_{\mathrm{o}}&\sim N(\boldsymbol{0},\Sigma_{11}\times\boldsymbol{A})\\ \boldsymbol{a}_{\mathrm{m}}|\boldsymbol{a}_{\mathrm{o}}&\sim N(\boldsymbol{a}_{\mathrm{o}}\tau,\Gamma\times\boldsymbol{A}),\end{aligned} \qquad (16)$$

with $\tau=\Sigma_{12}\Sigma_{11}^{-1}$ and $\Gamma=\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

Assume now the GIW density is used to represent prior uncertainty about the matrix $\boldsymbol{\Sigma}$, so that the Bartlett parameters (Σ_{11}, τ and Γ) can be regarded as being distributed

$$\begin{aligned}\Sigma_{11}&\sim S_0\chi^{-2}_{\upsilon_0}\\ \tau|\Gamma&\sim N(\tau_0,\Gamma\times H)\\ \Gamma&\sim S_1\chi^{-2}_{\upsilon_1+1},\end{aligned} \qquad (17)$$

where $S\chi^{-2}_{\upsilon}$ stands for a scaled inverted chi-square distribution with parameters (υ, S), a special case for an IW distribution with a scalar scale matrix. Note finally that the set of hyperparameters in Equation (17) is $\mathscr{H}=\{\upsilon_0,\upsilon_1,S_0,S_1,\tau_0,H\}$. All of these parameters must be defined by the analyst.

Conditional posterior distributions of the Bartlett parameters

As it is standard, the Bayesian analysis proceeds by forming the joint posterior distribution of all the unknowns that arises from the model. This is accomplished by multiplying the likelihood function times each of the prior densities. Next, the full conditional posterior distribution of any parameter of interest is derived by keeping the remaining ones fixed. In particular, the full conditional posterior distribution of the genetic covariance matrix $\boldsymbol{\Sigma}$ under the MAM model [Equation (12)] will be proportional to

$$\begin{aligned}p(\boldsymbol{\Sigma}|\mathscr{H},\mathscr{D})&\propto p(\boldsymbol{a}_{\mathrm{o}}|\Sigma_{11})\times p(\Sigma_{11}|S_0,\upsilon_0)\\ &\times p(\boldsymbol{a}_{\mathrm{m}}|\boldsymbol{a}_{\mathrm{o}},\tau,\Gamma)\times p(\tau|\Gamma,\tau_0,H)\times p(\Gamma|S_1,\upsilon_1),\end{aligned} \qquad (18)$$

where $\mathscr{D}=\left\{\boldsymbol{b},\boldsymbol{a}_{\mathrm{o}},\boldsymbol{a}_{\mathrm{m}},\boldsymbol{e}_{\mathrm{p}},\sigma^2_{e_{\mathrm{m}}},\sigma^2_{e_{\mathrm{o}}},\boldsymbol{y}\right\}$.

Explicitly, and after some rearrangement, we arrive at

$$\begin{aligned}&p(\boldsymbol{\Sigma}|\mathscr{H},\mathscr{D})\\ &\propto(\Sigma_{11})^{-\frac{1}{2}(q+\upsilon_0+2)}\times\exp\left\{-\frac{Q_{11}+S_0}{2\Sigma_{11}}\right\}\times(\Gamma)^{-\frac{1}{2}[(q+\upsilon_1+1)+2]}\\ &\times\exp\left\{-\frac{(\tau^2Q_{11}-2\tau Q_{12}+Q_{22})+H^{-1}(\tau-\tau_0)^2+S_1}{2\Gamma}\right\},\end{aligned} \qquad (19)$$

where we have made the use of the following notation for the symmetric matrix of sums of squares and cross-products

$$\boldsymbol{Q}=\begin{bmatrix}Q_{11} & Q_{12}\\ Q_{21} & Q_{22}\end{bmatrix}\equiv\begin{bmatrix}\boldsymbol{a}'_{\mathrm{o}}\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{o}} & \boldsymbol{a}'_{\mathrm{o}}\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{m}}\\ \boldsymbol{a}'_{\mathrm{m}}\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{o}} & \boldsymbol{a}'_{\mathrm{o}}\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{m}}\end{bmatrix}. \qquad (20)$$

Equation (19) evidences that the full conditional posterior distribution of the genetic covariance matrix $\boldsymbol{\Sigma}$ can be regarded as proportional to the product of three distinct distributions related to the Bartlett parameters, i.e.

$$p(\boldsymbol{\Sigma}|\mathscr{H},\mathscr{D}) \propto p(\Sigma_{11}|\mathscr{H},\mathscr{D}) \times p(\tau,\Gamma|\mathscr{H},\mathscr{D}) = p(\Sigma_{11}|\mathscr{H},\mathscr{D}) \times p(\tau|\Gamma,\mathscr{H},\mathscr{D}) \times p(\Gamma|\mathscr{H},\mathscr{D}). \quad (21)$$

We will show next that this density is in fact a GIW, by proving that these three distributions are in agreement with the ones corresponding to the Bartlett parameters as defined in Equation (10). The proof will be sketched here, as the focus is on presenting the main results. A more detailed derivation of some important results is left to the Appendix A.

Note first that the independence between Σ_{11} and the pair (τ,Γ) is self-evident. By keeping the factors that only depend on Σ_{11}, one can write

$$p(\Sigma_{11}|\mathscr{H},\mathscr{D}) \propto (\Sigma_{11})^{-\frac{1}{2}(q+\upsilon_0+2)} \times \exp\left\{-\frac{Q_{11}+S_0}{2\Sigma_{11}}\right\}. \quad (22)$$

On defining $\tilde{S}_0 = Q_{11} + S_0$ and $\tilde{\upsilon_0} = q + \upsilon_0$, Equation (22) is recognized as the kernel of the scaled inverted chi-square distribution

$$\Sigma_{11}|\mathscr{H},\mathscr{D} \sim \tilde{S}_0 \chi^{-2}_{\tilde{\upsilon_0}}. \quad (23)$$

Next, disregard all terms that do not depend on τ from the argument of the exponential function in Equation (19), so as to obtain

$$p(\tau|\Gamma,\mathscr{H},\mathscr{D}) \propto \exp\left\{-\frac{(\tau^2 Q_{11} - 2\tau Q_{12}) + H^{-1}(\tau-\tau_0)^2}{2\Gamma}\right\}. \quad (24)$$

After performing some algebraic manipulations on this expression, in Appendix A it is shown that

$$p(\tau|\Gamma,\mathscr{H},\mathscr{D}) \propto \exp\left\{-\frac{(Q_{11}+H^{-1})\left[\tau-(Q_{12}+\tau_0 H^{-1})(Q_{11}+H^{-1})^{-1}\right]^2}{2\Gamma}\right\}, \quad (25)$$

from where it is deduced that

$$\tau|\Gamma,\mathscr{H},\mathscr{D} \sim N\left(\frac{Q_{12}+\tau_0 H^{-1}}{Q_{11}+H^{-1}}, \frac{\Gamma}{Q_{11}+H^{-1}}\right). \quad (26)$$

Maybe a more insightful representation of the latter result can be achieved on defining the following identities (Brown 2002):

$$\tilde{H} \equiv \left(Q_{11}+H^{-1}\right)^{-1}, W \equiv \tilde{H} \times H^{-1} \quad \text{and} \quad \hat{\tau} \equiv Q_{11}^{-1} Q_{12}. \quad (27)$$

After some algebra, the parameters in Equation (26) can be expressed such that

$$\tau|\Gamma,\mathscr{H},\mathscr{D} \sim N\left(\tilde{\tau}_0, \Gamma \times \tilde{H}\right), \quad (28)$$

where $\tilde{\tau}_0 = W\tau_0 + (1-W)\hat{\tau}$. This representation indicates that the conditional posterior mean is a weighted average of the prior mean and of the information provided by the data through the quotient of quadratic forms. Note further that the weights will depend on the definition of the hyperparameter H. For instance, setting $H = S_0^{-1}$ is the standard choice if one is to retain the same mean structure as with the IW distribution (Brown 2002). In this case, $W = \left(Q_{11}S_0^{-1}+1\right)^{-1}$, which shows that if prior and data-based information on the direct additive variance are equal, so will be the weights given to both terms on the posterior mean of τ. On the other hand, as the information provided by the data increases, then the data-based term will be given a greater weight.

Continuing now with the main argument, it remains to be deduced the conditional posterior density of Γ. From Equation (19),

$$p(\Gamma|\mathscr{H},\mathscr{D}) \propto (\Gamma)^{-\frac{1}{2}[(q+\upsilon_1+1)+2]} \times \exp\left\{-\frac{S_1^* + S_1}{2\Gamma}\right\}, \quad (29)$$

where S_1^* is the result of collecting all terms from the exponential function that do not depend on τ. Equation (29) is easily recognized as the kernel of the following scaled inverted chi-square distribution

$$\Gamma|\mathscr{H},\mathscr{D} \sim \tilde{S}_1 \chi^{-2}_{\tilde{\upsilon_1}+1}, \quad (30)$$

with $\tilde{S}_1 = S_1^* + S_1$ and $\tilde{\upsilon_1} = q + \upsilon_1$. Further, it is shown in Appendix A that

$$S_1^* = (\boldsymbol{a}_m - \boldsymbol{a}_o\hat{\tau})'\boldsymbol{A}^{-1}(\boldsymbol{a}_m - \boldsymbol{a}_o\hat{\tau}) + WQ_{11}(\hat{\tau}-\tau_0)^2. \quad (31)$$

It is possible to gain some insight into the latter expression by noting that there are three sources of information contributing to the value that takes the scale parameter of the conditional distribution of Γ. First, there is prior information contributed by S_1. Second, there is information contributed by the data through the quadratic form on the adjusted maternal breeding values, as it can be seen in the first term in the right-hand side of Equation (31). In fact, it can be shown that this term is the expression for the

maximum likelihood estimator of Γ underthe distribution of $\boldsymbol{a}_m|\boldsymbol{a}_o$ if $\hat{\tau}$ is taken as the true value for τ. The third source of information apparently arises from the latter substitution, accounting for the fact that τ was estimated by $\hat{\tau}$.

Retrieving matrix $\boldsymbol{\Sigma}$

Equations (23), (28) and (30) imply that the genetic covariance matrix $\boldsymbol{\Sigma}$ follows conditionally a $GIW\left(\tilde{\upsilon}_0, \tilde{\upsilon}_1, \tilde{S}_0, \tilde{S}_1, \tilde{\tau}_0, \tilde{H}\right)$ distribution a posteriori. As a consequence, the conditional conjugacy property holds, and the estimation of the genetic (co)variance components can be attained through the Gibbs sampler. As a matter of fact, the algorithm simply requires sampling sequentially from the corresponding distributions of the Bartlett parameters and afterwards retrieving matrix $\boldsymbol{\Sigma}$ by applying the Bartlett decomposition backwards, i.e. by calculating

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{11}\tau \\ \tau\Sigma_{11} & \Gamma + \tau^2\Sigma_{11} \end{bmatrix}. \tag{32}$$

Different prior specifications

As we have seen, on assuming a GIW prior for the genetic covariance matrix Σ, the analyst must define the full set of hyperparameters $\mathscr{H} = \{\upsilon_0, \upsilon_1, S_0, S_1, \tau_0, H\}$, a task that bestows flexibility while specifying his prior knowledge. In the remainder of this section, we will discuss three different prior specifications. First, we will define a diffuse prior that reflects complete uncertainty about the covariance matrix a priori. Next, we will present the particular set $\mathscr{H}$ that equivalently retrieves a sample from an IW distribution. Finally, and based on this latter set, we will suggest to model differential uncertainty among scalar (co)variance components through a different specification of the parameters υ_0 and υ_1.

First assume that the prior distribution for $\boldsymbol{\Sigma}$ is diffuse with probability function proportional to $|\boldsymbol{\Sigma}|^{-\frac{1}{2}\upsilon} = (\Sigma_{11})^{-\frac{1}{2}\upsilon} \times \Gamma^{-\frac{1}{2}\upsilon}$, with $\upsilon = 3$ corresponding to the Jeffreys invariant prior, according to Brown (2002). Therefore, the full conditional posterior distribution of the genetic covariance matrix can be explicitly written as

$$\begin{aligned} p(\boldsymbol{\Sigma}|\mathscr{H}, \mathscr{D}) \propto (\Sigma_{11})^{-\frac{1}{2}(q+\upsilon)} \times \exp\left\{-\frac{Q_{11}}{2\Sigma_{11}}\right\} \\ \times (\Gamma)^{-\frac{1}{2}(q+\upsilon)} \times \exp\left\{-\frac{(\tau^2 Q_{11} - 2\tau Q_{12} + Q_{22})}{2\Gamma}\right\}. \end{aligned} \tag{33}$$

Now, resorting to the same arguments that we have used to derive the full conditional posterior distribution of the Bartlett parameters, it is verifiable that $\boldsymbol{\Sigma}$ follows conditionally a $GIW\left(\tilde{\upsilon}_0, \tilde{\upsilon}_1, \tilde{S}_0, \tilde{S}_1, \tilde{\tau}_0, \tilde{H}\right)$ distribution a posteriori, with $\tilde{\upsilon}_0 = q + \upsilon - 2$, $\tilde{\upsilon}_1 = q + \upsilon - 3$, and

$$\begin{aligned} \tilde{S}_0 &= Q_{11}, \\ \tilde{S}_1 &= Q_{22} - Q_{11}^{-1}Q_{12}^2, \\ \tilde{\tau}_0 &= Q_{11}^{-1}Q_{12}, \\ \tilde{H} &= Q_{11}^{-1}, \end{aligned} \tag{34}$$

where Q_{ij} symbolize the (i,j)-entry of the symmetric matrix of sums of squares and cross-products defined by Equation (20).

Conversely, assume an IW distribution prior for $\boldsymbol{\Sigma}$ under a conjugate approach, but consider the possibility of sampling sequentially from the conditional posterior distributions of the Bartlett parameters. This sampling strategy is advantageous from an algorithmic point of view, as it requires sampling only three standard normal deviates, while a straight forward generation will require many more (Smith & Hocking 1972). As a matter of fact, the available routines to sample from the Wishart distribution are based on this principle, as the algorithms usually rest on the Bartlett decomposition (e.g. the F77 WSHRT routine by Smith & Hocking 1972). It is shown in Appendix B that the equivalence is based on defining the following set of hyperparameters

$$\begin{aligned} S_0 &= \Sigma_{11}^*, \\ S_1 &= \Sigma_{22}^* - \Sigma_{12}^{*2}\Sigma_{11}^{*-1}, \\ \tau_0 &= \Sigma_{12}^*\Sigma_{11}^{*-1}, \\ H &= \Sigma_{11}^{*-1}, \end{aligned} \tag{35}$$

where $\boldsymbol{\Sigma}^* = \left\{\Sigma_{ij}^*\right\}$ represents the scale matrix of the prior distribution of $\boldsymbol{\Sigma}$, and further defining $\upsilon_0 = \upsilon + 1$ and $\upsilon_1 = \upsilon$, where υ is a common prior degree of belief parameter.

Indeed, a different specification of the parameters υ_0 and υ_1 can be used in a straightforward way to model the differential uncertainty expected between the estimates of the maternal and the direct additive variances. Such strategy is explored in the next section. In Appendix B, we present a sampling algorithm, easy to accommodate within the existing Gibbs sampling routines.

Prior specification using the GIW: an application

As a standard practice, most beef cattle breed associations run genetic evaluations on an annual or bian-

nual basis as a part of their performance recording programmes (BIF 2002, ch. 5). The main outcomes of these evaluations are breeding values predictions, or functions thereof, computed for sires, dams and their progeny in the population under study. The predictions, in turn, are obtained by solving the mixed model equations (cf. Henderson 1984) that arise from the model used to fit the data, conditional on the estimated (co)variance components. Now, although (co)variance components estimation should be performed before each run of the genetic evaluation, in practice, the estimation is usually undertaken every several years, when a large enough amount of data has been accrued. In any case, assume the estimation is accomplished through a hierarchical Bayesian analysis via the Gibbs sampler. In this context, it seems reasonable to use the posterior summaries of the distribution of the (co)variance components from the previous run to set the corresponding hyperparameters for the subsequent one, in the spirit of a Bayesian updating scheme. Taking advantage of the flexibility afforded by the GIW distribution, in this section, we describe the application of such strategy in a (co)variance components estimation problem using both field records and simulated data. The strategy is further compared against other prior specifications.

Field data sets

The full data set belongs to the firm 'Estancias y Cabañas Las Lilas', from Argentina, and comprised 7229 weaning weight records of Angus calves, born between 1972 and 2008. Every year, the firm undertakes a genetic evaluation as a selection tool and as a sales marketing strategy for their seedstock. Still, (co)variance components are not estimated with the same frequency, although several estimations were performed as the data have accrued over the years. In an attempt to mimic the scheme, we have used the full data file for creating two subsets. The first one included 4480 records from individuals born up to the year 1986, whereas the second subset contained all 6290 records taken on individuals born before the year 2000. A detailed description of the data sets we have analyzed is presented in Table 1.

The goal was to estimate (co)variance components via the Gibbs sampler for the full data file under several prior specifications. In general, prior distributions were parameterized as it is described by Sorensen & Gianola (2002, ch. 13.3). Specifically, scale parameters were set after specifying some 'reasonable' values as the mean of the prior distribution. In turn, the degrees of belief parameters were used to describe the uncertainty attached to those values. Beforehand, the MAM defined by Equation (12) was fitted, and (co)variance components were estimated using the ASReml (Gilmour *et al.* 2006) package [**REML**]. Next, several hierarchical Bayesian analyses were undertaken via the Gibbs sampler, with different strategies regarding prior opinion on the (co)variance components.

Table 1 Main features of the Angus and the simulated weaning weight data sets

	Subset1	Subset2	Full set	Simulated[1]
Records[2]	4480	6290	7229	4492
Pedigree	6080	8553	9936	5012
Pedigree connections[3]	5.9	14.7	21.9	–
Sires				
No	119	199	264	65
% of sires recorded	7	16	20	69
Mean number of calves	38	32	27	69
Dams				
No	1608	2127	2444	1376
% of dams recorded	45	55	57	64
Mean number of calves	3	3	3	3

[1]Averaged over 39 replicates. Some variability arose from an assumed fertility rate of 0.9.
[2]Weaning weights taken on individuals averaging 200 days of age.
[3]Non-zeroes in matrix A (in millions). Not computed for simulated data sets.

In the first strategy assayed, complete uncertainty was assumed, and thus, a diffuse prior sampling scheme was employed [**Diffuse**]. In this case, it was not necessary to define any hyperparameter. Instead, the estimation algorithm was based on sampling from fully conditional posterior distributions that depend only on functions of the quadratic forms of the data [see Equation (34)]. Furthermore, as the MCMC chains are independent of the initial values once the procedure has attained convergence, the coupling chain method (García-Cortés *et al.* 1998) could be used in a straight forward way to ascertain convergence.

Second, a meaningful prior opinion was considered through the knowledge of the REML point estimates, and then conjugated inverted-gamma distributions (inverted chi-square and IW) were assumed as priors for all the (co)variance components, parameterized so that they reflect uncertainty through the degrees of belief parameters. Specifically, the prior scale parameter for each (co)variance component was set after specifying the REML estimates as the mean of the prior distribution (see Appendix B for a detailed description of such param-

eterization). Further, aiming to evaluate the influence of this prior specification on the results, two other sets of prior mean values for the (co)variance components were used. These sets corresponded to the REML point estimates ± twice their standard errors. In turn, two different values for the degrees of belief parameters were set: 20 **[IW20]** and 100 **[IW100]**, respectively. These values were chosen to reflect, respectively, mild and strong confidence on the prior means. Table 2 contains a summary of the prior means and degrees of belief used in each of the analyses undertaken.

In the third strategy, an educated prior opinion was examined using the posterior summaries of the (co)variance components, obtained for each of the two data subsets through a diffuse sampling scheme estimation, to set the hyperparameters of the Bartlett prior densities in the full data set analysis (**[GIW_S$_1$]** and **[GIW_S$_2$]**, respectively). These analyses were undertaken assuming a GIW distribution for the genetic covariance matrix a priori. In such case, the full set $\mathscr{H}$ of hyperparameters needed to be defined. The specification was made in the following way. Using the outcomes from the subsets estimation, we first derived the degrees of belief parameters, υ_0 and υ_1, by equating the estimated marginal posterior means and variances of Σ_{11} and Γ to the theoretical means and variances of scaled inverted chi-square distributions, i.e.

$$\begin{aligned}\upsilon_0 &= \frac{2\times[\hat{m}_i(\Sigma_{11})]^2}{\hat{v}_i(\Sigma_{11})}+4,\\ \upsilon_1 &= \frac{2\times[\hat{m}_i(\Gamma)]^2}{\hat{v}_i(\Gamma)}+4,\end{aligned} \tag{36}$$

where $\hat{m}_i(\cdot)$ and $\hat{v}_i(\cdot)$ denote, respectively, the estimated marginal posterior mean and variance of the corresponding Bartlett parameter distributions, obtained from the data subset i (i = 1, 2). Next, we used the prior specification from Equation (35) to set the hyperparameters S_0, S_1, τ_0 and H. Specifically, the entries of the scale matrix $\boldsymbol{\Sigma}^*$ were computed as

$$\begin{aligned}\Sigma^*_{11} &= (\upsilon_0+2)\times\hat{M}_i(\Sigma_{11}),\\ \Sigma^*_{12} &= \Sigma^*_{21} = \Sigma^*_{11}\times\hat{M}_i(\tau),\\ \Sigma^*_{22} &= \left(\Sigma^{*2}_{12}\Sigma^{*-1}_{11}\right)+(\upsilon_1+3)\times\hat{M}_i(\Gamma),\end{aligned} \tag{37}$$

where $\hat{M}_i(\cdot)$ stands for the estimated marginal posterior mode of the corresponding Bartlett parameter distributions, obtained from data subset i (i = 1, 2). Note that this prior parameterization entails interpreting the estimated marginal posterior modes as some 'reasonable' values for the prior.

Table 2 Angus data. Prior means and degrees of belief used to specify prior knowledge for the different analyses undertaken in this study

	Prior means[1]			Degrees of belief[2]		
Analyses[3]	Dir. heritability	Mat. heritability	Dir–mat correl.	υ	υ_0	υ_1
REML	–	–	–	–	–	–
Diffuse	–	–	–	–	–	–
IW20_1	0.26	0.16	−0.69	20	–	–
IW20_2	0.20	0.12	−0.66	20	–	–
IW20_3	0.30	0.19	−0.75	20	–	–
IW100_1	0.26	0.16	−0.69	100	–	–
IW100_2	0.20	0.12	−0.66	100	–	–
IW100_3	0.30	0.19	−0.75	100	–	–
GIW_S$_1$	0.25	0.18	−0.71	–	32	34
GIW_S$_2$	0.21	0.14	−0.62	–	62	44
GIW_S$_2$|S$_1$	0.25	0.16	−0.67	–	105	85

[1]Figures in this table were computed as functions of the elicited prior means for the (co)variance components. The three sets of values in both IW20 and IW100 analyses correspond to the REML point estimates, REML − 2*SE and REML + 2*SE, respectively. For the different GIW analyses, in turn, prior means were defined as the (co)variance components posterior modes obtained after fitting the data.
[2]υ = degree of belief parameter from an inverted Wishart distribution; υ_0 and υ_1 are the degrees of belief parameters of the scaled inverted chi-square distributions from the Bartlett decomposition.
[3]Dashed lines subdivide analyses regarding uncertainty attached to the (co)variance components prior means into 'full uncertainty', 'mild prior opinion', 'strong prior opinion' and 'educated prior opinion' in ascending order.

A final analysis was launched using this latter strategy recursively, i.e. specifying the Bartlett parameters prior distributions for the second subset with the outcomes of the first one and then repeating the whole procedure for the full data set **[GIW_S$_2$|S$_1$]**. More precisely, assume that the firm undertook two different estimations of (co)variance components at the years 1986 and 2000. Assume further that in the 1986 evaluation, when data from subset 1 had been collected, the estimation was accomplished via the Gibbs sampler with a diffuse sampling scheme, as the analyst had no prior opinion on the values of the parameters at that time. At the year 2000, in turn, not only new data had been accrued (i.e. subset 2), but also the results from the previous estimation were available. Hence, the estimation procedure comprised fitting data subset 2 under a prior specification as the one described in the preceding paragraph. Likewise, the **GIW_S$_2$|S$_1$**

analysis involved using the posterior summaries obtained from this latter estimation to parameterize the prior distribution for the genetic covariance matrix before fitting the full data set. The values for the prior means and degrees of belief specified are displayed in Table 2.

Technical details concerning the implementation are described next. The Gibbs sampler used in this study was a single-site, systematic scan sampler, specifically written in Fortran 90. As a special feature, the genetic covariance matrix sampling step was coded following the algorithm presented in Appendix B. For every analysis, the programme was executed and a chain of 100 000 rounds was obtained. Further, the coupling method (García-Cortés *et al.* 1998) was implemented for the samples of the diffuse sampling scheme, and convergence was assessed considering a maximum difference between chains of 10^{-3} for every (co)variance component, which occurred at iteration 7606. Being conservative, thus, we discarded the first 10 000 samples in every outcome as burn-in. Posterior summaries for all (co)variance components and functions thereof, as well as autocorrelations among samples, were computed using the programme POSTGIBBSF90, from the BLUPF90 (Misztal *et al.* 2002) package. In particular, posterior means, posterior standard deviations, and autocorrelations for the direct heritability, the maternal heritability and the direct–maternal genetic correlation were used as the criteria for comparison.

Simulation study

For further insight into the estimation procedures, a stochastic simulation study was carried out. Closed random mating populations with overlapping generations were simulated. Within each replicate, a non-recorded base population of 500 cows was randomly mated to 20 bulls and produced the first generation of progeny. Phenotypes of these individuals were next sampled using the MAM [Equation (12)] as the data generation process. In the following step, the eldest sires and dams from the parental population were culled, according to a replacement rate of 0.25 for males and 0.20 for females. Their replacements were selected from the current generation using the estimated direct breeding values, obtained after fitting the MAM and solving the corresponding mixed model equations, as the selection criteria. A second generation was then created through random mating, with the proviso that parent–offspring matings be avoided, and the whole procedure was further repeated up to the tenth generation. Fifty such replicates were created and analyzed in this study. The main features of the simulated population structure, averaged over the replicates, were included in Table 1.

All simulated populations were analyzed using the same strategies regarding prior opinion about the (co)variance components as the ones employed with the Angus data set. For each replicate, the MAM was fitted, and (co)variance components were estimated via the **REML**, **Diffuse** and **IW100** analyses, as described earlier. In addition, a **GIW** analysis was undertaken in two steps. First, (co)variance components were estimated for a subset including records up to the eighth generation through a diffuse sampling scheme. Next, posterior summaries were computed and used to set the hyperparameters of the Bartlett prior densities in a full data set analysis. For

Table 3 Simulated data. Estimates and standard errors for direct heritability, maternal heritability and direct–maternal genetic correlation under different strategies with regard to prior opinion on the (co)variance components

	Direct heritability (True value = 0.25)		Maternal heritability (True value = 0.15)		Dir–mat correlation (True value = −0.70)	
Analyses[1]	Estimate	SE	Estimate	SE	Estimate	SE
REML	0.24 ± 0.04	0.05 ± 0.01	0.15 ± 0.03	0.04 ± 0.00	−0.69 ± 0.09	0.09 ± 0.02
Diffuse	0.24 ± 0.05	0.04 ± 0.01	0.14 ± 0.04	0.03 ± 0.01	−0.72 ± 0.13	0.09 ± 0.03
IW100_1	0.24 ± 0.04	0.03 ± 0.00	0.15 ± 0.03	0.02 ± 0.00	−0.69 ± 0.09	0.04 ± 0.01
IW100_2	0.17 ± 0.04	0.02 ± 0.00	0.08 ± 0.04	0.01 ± 0.00	−0.60 ± 0.16	0.06 ± 0.02
IW100_3	0.29 ± 0.05	0.03 ± 0.00	0.20 ± 0.03	0.02 ± 0.00	−0.72 ± 0.06	0.04 ± 0.01
GIW	0.23 ± 0.04	0.03 ± 0.00	0.14 ± 0.04	0.02 ± 0.01	−0.70 ± 0.12	0.06 ± 0.02

References: Estimate = REML point estimate or Bayesian posterior mean (averaged over 39 replicates ± standard deviation); SE = REML approximate standard error or Bayesian posterior standard deviation (averaged over 39 replicates ± standard deviation).

[1]Prior means for each replicate under the IW100 analyses correspond to the REML ± 2*SE estimates. Refer to main text for a detailed description of the GIW analysis.

Dashed lines subdivide analyses into ''full uncertainty'', strong prior opinion, and ''educated prior opinion'' in ascending order.

every Bayesian estimation procedure, 30 000 rounds of the Gibbs sampler were obtained: the first 10 000 were discarded as burn-in, and the remaining ones were used to compute posterior summaries. In the same way that we did with the field records data file, posterior means, posterior standard deviations, and autocorrelations for the direct heritability, the maternal heritability and the direct–maternal genetic correlation, averaged over replicates, were used to compare results.

A small digression is necessary at this point. Eleven of the fifty replicates have exhibited convergence problems while analyzing their corresponding subset. Basically, the marginal posterior modes of the (co)variance components taken as prior means for the full analyses have produced non-positive definite genetic covariance matrices, and thus, the Gibbs sampler has aborted. A deeper analysis showed that in those cases the coupling method failed to assess convergence before 10 000 rounds, which suggests a greater number of iterations would have been necessary. Trying not to obscure the conclusions then, and as the number of iterations per analysis was limiting because of run-time, we chose not to include those replicates. As a consequence, the results presented in this study were computed averaging the remaining 39 replicates.

Results

Estimates and standard errors of genetic parameters for the Angus weaning weight data are presented in Table 4. Focusing first on the estimates, note that even though not all analyses have returned exactly the same results, the figures were overall quite similar among procedures: direct and maternal heritabilities were in the order of 0.25 and 0.15, respectively, whereas the direct–maternal genetic correlation was around −0.69. Most variability was observed among the **IW100** analyses because of the strong influence exerted by the prior means we have set. On the contrary, the **IW20** analyses exhibited less dispersion and, indeed, the estimates were closer to the ones obtained under the **REML** and **Diffuse** analyses irrespective of the prior means. Now, regarding the standard errors, notice a consistent pattern has become apparent: in those analyses where a stronger weight has been put on the prior means, smaller standard errors were obtained compared with the less informative prior approaches. In particular, the $\mathbf{GIW\_S_2|S_1}$ recursive analysis showed the smallest standard errors.

In turn, Table 3 displays estimates and standard errors of genetic parameters for the simulated data sets, averaged over replicates. In general, the results followed the trend we have described for the analysis of field data. After fitting the same model we had used to generate the data, both the **REML** estimates and the **Diffuse** posterior means were on average very close to the true values simulated. Moreover, when the prior means of the genetic parameters in the **IW100** analyses were set to the corresponding REML estimates, the posterior means were also unbiased with respect to the true values. In addition,

Table 4 Angus data. Estimates and standard errors for direct heritability, maternal heritability and direct–maternal genetic correlation under different strategies with regard to prior opinion on the (co)variance components

	Direct heritability		Maternal heritability		Dir–mat correlation	
Analyses[1]	Estimate	SE	Estimate	SE	Estimate	SE
REML	0.26	0.04	0.16	0.03	−0.69	0.07
Diffuse	0.25	0.04	0.15	0.03	−0.69	0.08
IW20_1	0.25	0.04	0.15	0.02	−0.69	0.07
IW20_2	0.23	0.04	0.14	0.02	−0.69	0.07
IW20_3	0.26	0.04	0.16	0.02	−0.71	0.06
IW100_1	0.25	0.03	0.15	0.02	−0.69	0.04
IW100_2	0.21	0.03	0.12	0.02	−0.68	0.05
IW100_3	0.29	0.03	0.18	0.02	−0.73	0.04
GIW_S_1	0.28	0.04	0.17	0.02	−0.71	0.05
GIW_S_2	0.24	0.03	0.15	0.02	−0.66	0.05
$GIW\_S_2\|S_1$	0.27	0.02	0.16	0.02	−0.69	0.04

References: Estimate = REML point estimate or Bayesian posterior mean; SE = REML approximate standard error or Bayesian posterior standard deviation.

[1]Refer to main text and Table 2 for a full description of the analyses undertaken. Dashed lines subdivide analyses into 'full uncertainty', ''mild prior opinion'', ''strong prior opinion'' and ''educated prior opinion'' in ascending order.

standard errors were smaller. Likewise, the **GIW** analysis returned on average accurate estimates and reduced standard errors. Now, when the prior means were set to over-dispersed values in the **IW100** analyses, posterior means deviated from the true genetic parameters.

Figure 1 shows autocorrelation plots of genetic parameters for the Angus data set. For a better representation, only one of the three correlograms for the **IW20** and **IW100** analyses has been plotted as the curves within analysis were very similar. Likewise, we have depicted only the **GIW_$S_2|S_1$** autocorrelation plot. Notice that the **IW100** and **GIW** analyses showed better convergence rates compared with the **Diffuse** and the **IW20** analyses.

Finally, lag10 and lag200 autocorrelations of the genetic parameters for the simulated data sets are presented in Table 5. Again, an improved mixing of the chain and thus faster convergence were observed when a stronger weight was placed on the prior distributions. In fact, **IW100** and **GIW** analyses showed the better convergence behaviour. However, differences in the autocorrelations between these two analyses have arisen. Looking at lag200 autocorrelations, it becomes apparent that those differences are more important for the maternal heritability and the direct–maternal genetic correlation than for the direct heritability. In this respect, it is worth recalling that in the **IW100** analyses, the whole prior genetic covariance matrix distribution was constrained by a single degree of belief parameter ($\upsilon = 100$), whereas in the **GIW** analyses, two different degrees of belief parameters were involved (averaging $\upsilon_0 = 64$ and $\upsilon_1 = 19$, respectively).

Discussion

In this study, the GIW distribution is introduced into the animal breeding literature. The description was based extensively on the works by Brown (2002) and Le *et al.* (1999). In particular, we have acknowledged its flexibility to elicit prior knowledge regarding the distribution of the genetic covariance matrix in the framework of a hierarchical Bayesian analysis. Still, other applications may arise following the reasoning described here. For instance, a straight forward application would be to use the GIW distribution as it was originally intended for, i.e. as a

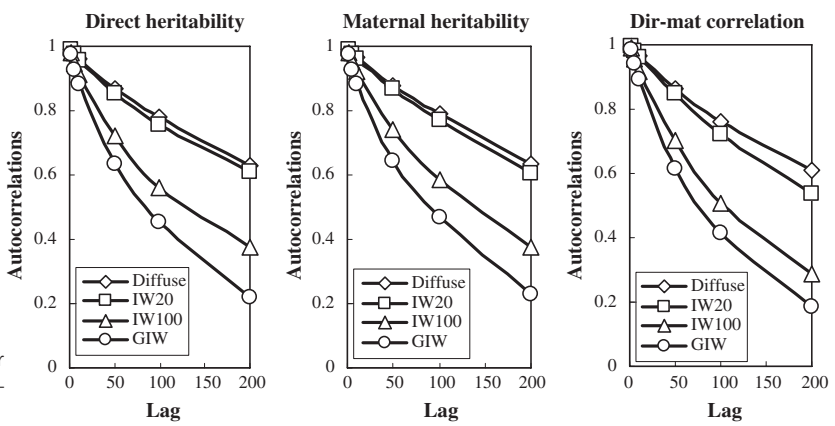

Figure 1 Angus data. Autocorrelation plots for direct heritability, maternal heritability and direct–maternal genetic correlation.

Table 5 Simulated data. Lag10 and lag200 autocorrelations among samples for direct heritability, maternal heritability and direct–maternal genetic correlation

	Direct heritability		Maternal heritability		Dir–mat correlation	
Analyses[1]	Lag10	Lag200	Lag10	Lag200	Lag10	Lag200
Diffuse	0.92 ± 0.06	0.48 ± 0.14	0.95 ± 0.06	0.62 ± 0.14	0.96 ± 0.02	0.55 ± 0.17
IW100_1	0.83 ± 0.03	0.12 ± 0.07	0.84 ± 0.02	0.12 ± 0.08	0.83 ± 0.02	0.09 ± 0.06
IW100_2	0.83 ± 0.03	0.15 ± 0.09	0.87 ± 0.02	0.20 ± 0.12	0.85 ± 0.02	0.15 ± 0.07
IW100_3	0.82 ± 0.03	0.11 ± 0.06	0.82 ± 0.02	0.10 ± 0.08	0.80 ± 0.02	0.07 ± 0.05
GIW	0.85 ± 0.03	0.19 ± 0.10	0.91 ± 0.04	0.33 ± 0.16	0.90 ± 0.04	0.29 ± 0.23

References: Lag10 and lag 200 autocorrelations averaged over 39 replicates (± SD).

[1]Prior means for each replicate under the IW100 analyses correspond to the REML ± 2*SE estimates. Refer to main text for a detailed description of the GIW analysis. Dashed lines subdivide analyses into 'full uncertainty', 'strong prior opinion' and 'educated prior opinion' in ascending order.

natural specification for the prior covariance structure of multivariate Gaussian data with a monotone pattern of missing data (Garthwaite & Al-Awadhi 2001). Such an application in the context of multiple-trait animal models is the core of the full conjugate Gibbs sampler (FCG), an alternative procedure to 'data augmentation' algorithms with faster convergence rates (Cantet *et al.* 2004).

Additionally, we have derived theoretical results regarding the use of the GIW as the prior distribution for the genetic covariance matrix of the MAM in the context of a hierarchical Bayesian analysis. In particular, we have proven that the conditional conjugacy property holds for the GIW, and hence, (co)variance component estimation through a Gibbs sampler using this distribution is feasible. In fact, it has been shown that the IW can be regarded as a special case of the GIW, by setting a specific set of hyperparameters. Further, we have demonstrated that both an extension to represent differential uncertainty and a diffuse prior specification are straightforward. Finally, we have sketched a sampling algorithm easy to fit within existing Gibbs sampler routines.

Now, eliciting priors is another issue. In this research, we have studied a strategy that we believe arise naturally given the standard practice of genetic evaluations. The strategy is based on using previously estimated values of the (co)variance components to assess the hyperparameters on the next round, in a recursive fashion. In particular, we have derived the degrees of belief of the Bartlett parameters by equating marginal posterior means and variances from a subset of the data to their corresponding theoretical expectation and variance. Thus, we have followed an intuitive Bayesian updating approach, exploiting 'the ''memory'' property of the Bayes theorem' (Gianola & Fernando 1986). A more formal treatment is lacking, however, as the strategy poses a question we have not answered in this study: does this recursive procedure truly 'target and pursue' the unknown parameters, or does it get stuck with the singularity of the estimates that may occur in a data subset? The answer may lay in the fact that, ultimately, the genetic parameters are not immutable quantities over time, as their true values may be redefined as information accrues and data structure changes.

In any case, our proposal was tested on both field records and simulated data and further compared with other prior specification approaches. The recursive strategy has returned accurate point estimates and reduced standard errors when compared with non-informative prior settings while improving considerably the convergence rates. A note of caution is in order here, as these advantages have appeared to be associated with the value of the prior degrees of belief specified. In fact, the IW analyses with strong prior opinion have also produced low standard errors and better convergence behaviour. However, when the prior means were set to over-dispersed values, the estimates were biased with respect to the true values simulated. The risk of biasing the estimates by using strong prior opinion has already been pointed out in the review by Misztal (2008).

In conclusion, we have shown that differential uncertainty regarding prior knowledge on the genetic (co)variance components in the framework of a MAM is easily accounted for through a GIW prior specification for the genetic covariance matrix. Moreover, as conditional conjugacy holds, parameter estimation can be readily accomplished via the Gibbs sampler.

Acknowledgements

The authors thank Dr. Claudio Fioretti (Estancias y Cabaña Las Lilas S.A., Argentina) for kindly providing the data used in this study. Hugo Von Bernard and María José Suárez have worked hard on editing this data set. Two anonymous reviewers contributed to improve the description of the application study we have undertaken. Funding for this research was provided by grants of CONICET (PIP 0833/10), Secretaría de Ciencia y Técnica, UBA (UBACyT G042/08) and Agencia Nacional de Ciencia y Tecnología (PICT 1863/06), of Argentina.

References

Bartlett M.S. (1933) On the theory of statistical regression. *Proc. R. Soc. Edinb.*, **53**, 260–283.

Bauwens L., Lubrano M., Richard J.F. (1999) Bayesian Inference in Dynamic Econometric Models. Oxford University Press, New York, USA.

Beef Improvement Federation (BIF) (2002) Guidelines for Uniform Beef Improvement Programs, 8th edn. Univ. of Georgia, Athens, USA.

Brown P.J. (2002) The generalized inverted Wishart distribution. In: A.H. El-Shaarawi, W.W. Piegorsch (eds), Encyclopaedia of Environmetrics. Wiley, England, pp. 1079–1083.

Brown P.J., Le N.D., Zidek J.V. (1994) Inference for a covariance matrix. In: P.R. Freeman, A.F.M. Smith (eds), Aspects of Uncertainty: A Tribute to D. V. Lindley. Wiley, New York, pp. 77–92.

Cantet R.J.C., Fernando R.L., Gianola D. (1992) Bayesian inference about dispersion parameters of univariate

mixed models with maternal effects: theoretical considerations. *Genet. Sel. Evol.*, **24**, 107–135.
Cantet R.J.C., Birchmeier A.N., Steibel J.P. (2004) Full conjugate analysis of normal multiple traits with missing records using a generalized inverted Wishart distribution. *Genet. Sel. Evol.*, **36**, 49–64.
Daniels M.J., Pourahmadi M. (2002) Bayesian analysis of covariance matrices and dynamic models for longitudinal data. *Biometrika*, **89**, 553–566.
García-Cortés L.A., Rico M., Groeneveld E. (1998) Using coupling with the Gibbs sampler to assess convergence in animal models. *J. Anim. Sci.*, **76**, 441–447.
Garthwaite P.H., Al-Awadhi S.A. (2001) Non-conjugate prior distribution assessment for multivariate normal sampling. *J. R. Stat. Soc. Ser. B*, **63**, 95–110.
Gelfand A.E., Smith A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.*, **85**, 398–409.
Gianola D., Fernando R.L. (1986) Bayesian methods in animal breeding theory. *J. Anim. Sci.*, **63**, 217–244.
Gilmour A.R., Gogel B.J., Cullis B.R., Thompson R. (2006) ASReml User Guide Release 2.0. VSN International Ltd, Hemel Hempstead HP1 1ES, UK.
Henderson C.R. (1984) Applications of Linear Models in Animal Breeding. University of Guelph, Guelph, Ontario, Canada.
Jensen J., Wang C.S., Sorensen D.A., Gianola D. (1994) Bayesian inference on variance and covariance components for traits influenced by maternal and direct genetic effects, using the Gibbs sampler. *Acta Agri. Scand. A-An.*, **44**, 193–201.
Le N.D., Zidek J.V. (2006) Statistical Analysis of Environmental Space-Time Processes. Springer Science+Business Media, Inc., NY, USA.
Le N.D., Sun L., Zidek J.V. (1999) Bayesian Spatial Interpolation and Backcasting Using Gaussian – Generalized Inverted Wishart Model. Technical Report 185. Statistics Dept., UBC, Vancouver, Canada.
Misztal I. (2008) Reliable computing in estimation of variance components. *J. Anim. Breed. Genet.*, **125**, 363–370.
Misztal I., Tsuruta S., Strabel T., Auvray B., Druet T., Lee D.H. (2002) BLUPF90 and related programs (BGF90). In: 7th World Congress on Genetics Applied to Livestock Production, Montpellier, 19–23 August 2002.
Quaas R., Pollak E. (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *J. Anim. Sci.*, **51**, 1277–1287.
Searle S.R. (1982) Matrix Algebra Useful for Statistics. John Wiley & Sons, NY, USA.
Smith W.B., Hocking R.R. (1972) Algorithm AS 53: Wishart variate generator. *Appl. Stat.*, **21**, 341–345.
Sorensen D.A., Gianola D. (2002) Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics. Springer-Verlag, London, UK.
Willham R.L. (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics*, **19**, 18–27.

Appendices

Appendix A. Results on posterior conditional distributions

Several results were presented in connection with the full conditional distributions of the Bartlett parameters resulting from the decomposition of the genetic covariance matrix $\boldsymbol{\Sigma}$ under the MAM [Equation (12)]. Here a more detailed derivation of those results is provided.

Let us start with the full conditional distribution of τ as expressed in Equation (24), arrived at after disregarding all terms that do not depend on τ from the argument of the exponential function in the Equation (19)

$$p(\tau|\Gamma,\mathscr{H},\mathscr{D}) \propto \exp\left\{-\frac{(\tau^2 Q_{11} - 2\tau Q_{12}) + H^{-1}(\tau-\tau_0)^2}{2\Gamma}\right\}. \tag{A.1}$$

Completing the squares in the first term in the exponential produces

$$\begin{aligned}\tau^2 Q_{11} - 2\tau Q_{12} &= Q_{11}\left(\tau^2 - 2\tau Q_{11}^{-1}Q_{12}\right)\\ &= Q_{11}\left[\tau^2 - 2\tau Q_{11}^{-1}Q_{12} + \left(Q_{11}^{-1}Q_{12}\right)^2 - \left(Q_{11}^{-1}Q_{12}\right)^2\right]\\ &= Q_{11}\left(\tau - Q_{11}^{-1}Q_{12}\right)^2 - Q_{11}^{-1}Q_{12}^2.\end{aligned} \tag{A.2}$$

The last term in Equation (A.2) does no depend on τ, and thus, it is absorbed within the normalizing constant. The next step involves combining the quadratic forms.

$$Q_{11}\left(\tau - Q_{11}^{-1}Q_{12}\right)^2 + H^{-1}(\tau-\tau_0)^2 \tag{A.3}$$

To do so, we make use of the following identity (Sorensen & Gianola 2002, p. 227)

$$\begin{aligned}M(z-m)^2 + B(z-b)^2 &= (M+B)(z-c)^2\\ &\quad + \frac{MB}{M+B}(m-b)^2,\end{aligned} \tag{A.4}$$

with $c = (M+B)^{-1}(Mm+Bb)$. Note that on equating $M = Q_{11}$, $m = Q_{12}Q_{11}^{-1}$, $B = H^{-1}$, $b = \tau_0$, and $z = \tau$, i.e. subsequently dropping the second term in Equation (A.4), as it does not depend on τ, we arrive

at Equation (25), from where it can be deduced that a posteriori τ is conditionally distributed as a univariate normal variable.

Next, we will derive the expression for the scale parameter of the conditional posterior distribution of Γ, i.e. $\tilde{S}_1$, as presented in Equation (31). As it was stated before, this step involves collecting all factors from the exponential function in Equation (19) that do not depend on τ. These are (i) the prior scale parameter for Γ, i.e. S_1 ; (ii) the quadratic form in the maternal breeding values, Q_{22}; (iii) the term $-Q_{11}^{-1}Q_{12}^2$, discarded while completing squares in Equation (A.2); and (iv) the second term in the right-hand side of Equation (A.4), which after the appropriate replacement and on using the set of identities defined in Equation (27), renders $WQ_{11}(\hat{\tau}-\tau_0)^2$.

Now, operating on the quadratic forms by adding and subtracting $Q_{11}^{-1}Q_{12}^2$,

$$\begin{aligned} Q_{22}-2Q_{11}^{-1}Q_{12}^2+Q_{11}^{-1}Q_{12}^2 &= Q_{22}-2Q_{12}\hat{\tau}+Q_{12}\hat{\tau} \\ &= Q_{22}-2Q_{12}\hat{\tau}+Q_{11}\hat{\tau}^2 \\ &= \boldsymbol{a}_{\mathrm{m}}'\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{m}}-2\left(\boldsymbol{a}_{\mathrm{o}}'\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{m}}\right)\hat{\tau} \\ &\quad +\boldsymbol{a}_{\mathrm{o}}'\boldsymbol{A}^{-1}\boldsymbol{a}_{\mathrm{o}}\hat{\tau}^2 \\ &= (\boldsymbol{a}_{\mathrm{m}}-\boldsymbol{a}_{\mathrm{o}}\hat{\tau})'\boldsymbol{A}^{-1}(\boldsymbol{a}_{\mathrm{m}}-\boldsymbol{a}_{\mathrm{o}}\hat{\tau}), \end{aligned} \tag{A.5}$$

and thus

$$\tilde{S}_1=(\boldsymbol{a}_{\mathrm{m}}-\boldsymbol{a}_{\mathrm{o}}\hat{\tau})'\boldsymbol{A}^{-1}(\boldsymbol{a}_{\mathrm{m}}-\boldsymbol{a}_{\mathrm{o}}\hat{\tau})+WQ_{11}(\hat{\tau}-\tau_0)^2+S_1. \tag{A.6}$$

Formulas regarding the conditional posterior distribution of a covariance matrix in the more general case of multiple-block Bartlett decomposition can be found in Brown (2002) and Le *et al.* (1999).

Appendix B. Results on prior specification and the sampling algorithm

Assume now that the analyst defines an IW prior for $\boldsymbol{\Sigma}$ under a conditional conjugate approach. In such case, we have stated that Equation (35) defines the particular set of prior hyperparameters of a GIW distribution that equivalently retrieves a sample from the corresponding conditional posterior IW distribution. Here, we prove such equivalence.

More specifically, consider a Bayesian analysis of the MAM via the Gibbs sampler as described in Sorensen & Gianola (2002). There, the genetic covariance matrix is sampled from the conditional posterior IW distribution defined by $IW(\upsilon+q,\boldsymbol{Q}^*)$, with $\boldsymbol{Q}^*=\boldsymbol{Q}+\boldsymbol{\Sigma}^*$. Furthermore, $\boldsymbol{Q}$ is the matrix of sums of squares and cross-products defined in Equation (20), and $\boldsymbol{\Sigma}^*$ represents the scale matrix of the prior distribution of $\boldsymbol{\Sigma}$, usually parameterized as $\boldsymbol{\Sigma}^*=\upsilon\boldsymbol{S}^*$, where $\boldsymbol{S}^*$ represents a matrix of a priori 'reasonable' values for the genetic (co)variance components, and υ is a common degree of belief on those values a priori.

On the other hand, it can be shown that replacing the set of hyperparameters in Equation (35) in Equations (23), (28) and (30) yields the following posterior conditional distributions for the Bartlett parameters

$$\begin{aligned} \Sigma_{11}|\mathscr{H},\mathscr{D} &\sim Q_{11}^*\chi_{\tilde{\upsilon}_0}^{-2}, \\ \tau|\Gamma,\mathscr{H},\mathscr{D} &\sim N\left(Q_{11}^{*-1}Q_{12}^*,Q_{11}^{*-1}\Gamma\right), \\ \Gamma|\mathscr{H},\mathscr{D} &\sim \left(Q_{22}^*-Q_{11}^{*-1}Q_{12}^{*2}\right)\chi_{\tilde{\upsilon}_1+1}^{-2}, \end{aligned} \tag{B.1}$$

with $\tilde{\upsilon}_0=\upsilon_0+q=\upsilon+q+1$ and $\tilde{\upsilon}_1=\upsilon_1+q=\upsilon+q$.

Now, to prove that both sampling strategies are equivalent, it will be shown next that the product of the kernels of the three densities in Equation (B.1) retrieves the kernel of the appropriate IW distribution. Explicitly, and after replacing the Bartlett parameters with the corresponding entries of the covariance matrix $\boldsymbol{\Sigma}$, the multiplication yields

$$\begin{aligned} &p(\Sigma_{11}|\mathscr{H},\mathscr{D})\times p(\tau|\Gamma,\mathscr{H},\mathscr{D})\times p(\Gamma|\mathscr{H},\mathscr{D}) \\ &\quad\propto (\Sigma_{11})^{-\frac{1}{2}[(\upsilon+q+1)+2]}\times(\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2)^{-\frac{1}{2}[(\upsilon+q)+1+2]} \\ &\quad\times\exp\left\{-\frac{1}{2}\left[\frac{Q_{11}^*}{\Sigma_{11}}+\frac{Q_{11}^*(\Sigma_{11}^{-1}\Sigma_{12}-Q_{11}^{*-1}Q_{12}^*)^2}{(\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2)}\right.\right. \\ &\quad\left.\left.+\frac{(Q_{22}^*-Q_{11}^{*-1}Q_{12}^{*2})}{(\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2)}\right]\right\}. \end{aligned} \tag{B.2}$$

Note first that

$$\begin{aligned} &(\Sigma_{11})^{-\frac{1}{2}[(\upsilon+q+1)+2]}\left(\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2\right)^{-\frac{1}{2}[(\upsilon+q)+1+2]} \\ &\quad=(\Sigma_{11}\Sigma_{22}-\Sigma_{21}\Sigma_{12})^{-\frac{1}{2}(\upsilon+q+3)}=|\boldsymbol{\Sigma}|^{-\frac{1}{2}(\upsilon+q+3)}. \end{aligned} \tag{B.3}$$

Now, working out the exponentials, we arrive at

$$\begin{aligned}
&\exp\left\{-\frac{1}{2}\left[\frac{Q_{11}^*}{\Sigma_{11}}+\frac{Q_{11}^*\left(\Sigma_{11}^{-1}\Sigma_{12}-Q_{11}^{*-1}Q_{12}^*\right)^2+\left(Q_{22}^*-Q_{11}^{*-1}Q_{12}^{*2}\right)}{\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2}\right]\right\}\\
&\quad=\exp\left\{-\frac{1}{2}\left[\frac{Q_{11}^*}{\Sigma_{11}}+\frac{Q_{11}^*\Sigma_{11}^{-2}\Sigma_{12}^2-2Q_{12}^*\Sigma_{11}^{-1}\Sigma_{12}+Q_{22}^*}{\Sigma_{22}-\Sigma_{11}^{-1}\Sigma_{12}^2}\right]\right\}\\
&\quad=\exp\left\{-\frac{1}{2}\left[\left(\frac{Q_{11}^*\Sigma_{22}-Q_{12}^*\Sigma_{12}}{|\boldsymbol{\Sigma}|}\right)+\left(\frac{Q_{22}^*\Sigma_{11}-Q_{12}^*\Sigma_{12}}{|\boldsymbol{\Sigma}|}\right)\right]\right\}\\
&\quad=\exp\left\{-\frac{1}{2}\left[\left(Q_{11}^*\Sigma^{11}+Q_{12}^*\Sigma^{12}\right)+\left(Q_{22}^*\Sigma^{22}-Q_{12}^*\Sigma^{12}\right)\right]\right\}\\
&\quad=\exp\left\{-\frac{1}{2}tr\left(\Sigma^{-1}\boldsymbol{Q}^*\right)\right\},
\end{aligned}\tag{B.4}$$

where Σ^{ij} symbolize the (i,j)-entry of matrix Σ^{-1}.

Equations (B.3) and (B.4) show that Equation (B.2) can be recognized as the kernel of an $IW(\upsilon+q,\boldsymbol{Q}^*)$ distribution. Therefore, an IW prior for the genetic covariance matrix under a MAM can be regarded as a special case for a GIW prior when defining the specific set of hyperparameters in Equation (35). Moreover, note that Equation (B.1) suggests a sampling algorithm. Assume the analyst wishes to set different values for the parameters υ_0 and υ_1 to reflect differential uncertainty a priori. Then, the following algorithm will retrieve samples from the conditional posterior distribution of the genetic covariance matrix $\boldsymbol{\Sigma}$:

1 (a) Define υ_0 and υ_1, and form matrix $\boldsymbol{\Sigma}^*$ sequentially with the following entries:

$$\Sigma_{11}^*=(\upsilon_0+2)S_{11}^*.$$

$$\Sigma_{12}^*=\Sigma_{21}^*=\left(S_{12}^*S_{11}^{*-1}\right)\Sigma_{11}^*.$$

$$\Sigma_{22}^*=(\upsilon_1+3)\left(S_{22}^*-S_{12}^{*2}S_{11}^{*-1}\right)+\left(\Sigma_{12}^{*2}\Sigma_{11}^{*-1}\right).$$

(b) Compute $\tilde{\upsilon}_0=q+\upsilon_0$ and $\tilde{\upsilon}_1=q+\upsilon_1$.

2 Form matrix $\boldsymbol{Q}^*=\boldsymbol{Q}+\boldsymbol{\Sigma}^*$.

3 Sample Σ_{11} from $\Sigma_{11}|\mathscr{H},\mathscr{D}\sim Q_{11}^*\chi_{\tilde{\upsilon}_0}^{-2}$.

4 Sample Γ from $\Gamma|\mathscr{H},\mathscr{D}\sim\left(Q_{22}^*-Q_{11}^{*-1}Q_{12}^{*2}\right)\chi_{\tilde{\upsilon}_1+1}^{-2}$.

5 Sample τ from $\tau|\Gamma,\mathscr{H},\mathscr{D}\sim N\left(Q_{11}^{*-1}Q_{12}^*,Q_{11}^{*-1}\Gamma\right)$.

6 Retrieve matrix $\boldsymbol{\Sigma}$ by calculating

$$\boldsymbol{\Sigma}=\begin{bmatrix}\Sigma_{11} & \Sigma_{11}\tau\\ \tau\Sigma_{11} & \Gamma+\tau^2\Sigma_{11}\end{bmatrix}.$$

7 Repeat 2–6 within each cycle of the Gibbs sampler.

The definition of the prior scale matrix in the step 1(a) of the algorithm aforementioned is based on using some 'reasonable' values for the prior genetic (co)variance components (the entries in matrix $\boldsymbol{S}^*$) as statements about the mode of the distributions of the Bartlett parameters and next on solving back for each entry of matrix $\boldsymbol{\Sigma}^*$. In particular, on defining $\upsilon_0=\upsilon+1$ and $\upsilon_1=\upsilon$, the algorithm will retrieve an IW sample, with prior scale matrix $\boldsymbol{\Sigma}^*=(\upsilon+3)\boldsymbol{S}^*$. In that case, matrix $\boldsymbol{S}^*$ represents a statement about the mode of the corresponding IW prior distribution.