THE GENERALIZED INVERTED WISHART DISTRIBUTION

Philip J. Brown
University of Kent at Canterbury, UK

The *generalized inverted Wishart (GIW) distribution* is a flexible extension of the inverted Wishart (IW) distribution which overcomes the major deficiency of that distribution whilst retaining much of the ease of computation inherent in the natural conjugate roots of the inverted Wishart distribution in Bayesian analysis for normal sampling.

In offering a more flexible prior distribution for the covariance matrix of a multivariate normal distribution it has general applicability. It also offers a conjugate Bayesian analysis for problems of missing data where the missing data has a monotone structure. It has been especially useful in simultaneous modelling of the spatial distribution of ozone where monitoring stations come on stream in sequence, using multivariate regression analysis for interpolation and prediction, (Le *et al.*, 1999).

The inverted Wishart distribution, denoted $\mathcal{IW}(\delta, \boldsymbol{G})$, is the natural conjugate prior distribution for the $q \times q$ covariance matrix $\boldsymbol{\Sigma}$ in multivariate normal sampling from a $q-$variate normal distribution mean zero and covariance matrix $\boldsymbol{\Sigma}$. It has one major deficiency: whilst its expectation has a full complement of $q(q+1)/2$ prior parameters to specify through the symmetric positive definite $q \times q$ matrix $\boldsymbol{G}$, the uncertainty attached to this matrix is governed by the single scalar parameter δ.

We will present the generalized inverted Wishart in its full generality and then specialize and motivate it, giving some properties in the process. The distribution was introduced by Brown *et al* (1994a). An earlier special case is the 'recursive extended natural conjugate' of Richard and Steel (1988). The idea of recursive extension has also been used in connection with the multinomial through a sequence of binomials (Connor and Mosimann, 1969).

1 Characterising the generalized inverted Wishart

The definition of the GIW requires a partition of the q parameters into $k \geq 1$ blocks. Let $\{q_i\}_{i=1}^k, q_i > 0$ be the partition and $q = q_1 + q_2 + \ldots, q_k$. The specification now takes a chain of conditional distributions. We will take the sequence to be ordered as the indices from 1 (first) to k (last). Define the cumulative sums

$$q_{\{i\}} = q_1 + \ldots + q_i.$$

We will be concerned with the sequence: the distribution of the first set of q_1 variables, next the distribution of the the second set of q_2 variables given the first q_1 variables, next the distribution of the q_3 third partition variables given all the previous variables, $q_{\{2\}}$ in number, down to the kth set of variables given all the previous variables, $q_{\{k-1\}}$ in number. The overall distribution is obtained as the product of these conditional distributions. We partition the covariance matrix $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{ij})$ conformally where $\boldsymbol{\Sigma}_{ij}$ is a $q_i \times q_j$ submatrix.

Also using the $\{.\}$ notation as above for cumulative indices we define the sequences of 'slopes' $\boldsymbol{B}_i$ and residual covariance matrices by

$$\boldsymbol{B}_i = \boldsymbol{\Sigma}_{\{i\}\{i\}}^{-1}\boldsymbol{\Sigma}_{\{i\}(i+1)} \tag{1}$$

$$\boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}_{(i+1)(i+1)} - \boldsymbol{\Sigma}_{(i+1)\{i\}}\boldsymbol{\Sigma}_{\{i\}\{i\}}^{-1}\boldsymbol{\Sigma}_{\{i\}(i+1)} \tag{2}$$

for $i = 1, \ldots, (k-1)$.

We adopt the matrix-variate notation of Dawid (1981). If $\boldsymbol{U}(p\times p)$ is inverted Wishart with shape parameter δ and scale matrix $\boldsymbol{F}$ then for $\delta > 0$

$$\boldsymbol{U} \sim \mathcal{IW}(\delta, \boldsymbol{F}),$$

by which we mean $\boldsymbol{U}^{-1}$ has a Wishart distribution with scale matrix $\boldsymbol{F}^{-1}$ and degrees of freedom $(\delta+p-1)$. We have $E(\boldsymbol{U}) = \boldsymbol{F}/(\delta-2)$ for $\delta > 2$. This notation preserves the shape parameter under marginalization. The notation for a matrix- normal distribution is also required. Suppose $\boldsymbol{V}$ is a $r\times p$ matrix of independent standard normal entries. Also let $\boldsymbol{M}, \boldsymbol{A}, \boldsymbol{C}$ be matrices of constants of size $r\times p, r\times r, p\times p$, respectively. Then $\boldsymbol{W} = \boldsymbol{M} + \boldsymbol{A}\boldsymbol{V}\boldsymbol{C}'$ is said to have a matrix-normal distribution, $\boldsymbol{W} - \boldsymbol{M} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{A}', \boldsymbol{C}\boldsymbol{C}')$, so that the first argument of $\mathcal{N}(.,.)$ is proportional to the covariances of columns, the second that of rows. This notation preserves the matrix structure and avoids stringing out matrices as vectors and resorting to Kronecker products.

Using this notation we can set up the sequence of submodules described above. Suppose the row vector $\boldsymbol{Y} \sim \mathcal{N}(1,\boldsymbol{\Sigma})$ then the model formed from successive partitions is

$$\begin{aligned}
\boldsymbol{Y}_1 &\sim \mathcal{N}(1, \boldsymbol{\Sigma}_{11}) \\
\boldsymbol{Y}_2 \mid \boldsymbol{Y}_1 &\sim \boldsymbol{Y}_1\boldsymbol{B}_1 + \mathcal{N}(1, \boldsymbol{\Sigma}_{22.1}) \\
\boldsymbol{Y}_3 \mid \boldsymbol{Y}_{\{2\}} &\sim \boldsymbol{Y}_{\{2\}}\boldsymbol{B}_2 + \mathcal{N}(1, \boldsymbol{\Sigma}_{33.\{2\}}) \\
\vdots \quad &\vdots \quad \vdots \\
\boldsymbol{Y}_k \mid \boldsymbol{Y}_{\{k-1\}} &\sim \boldsymbol{Y}_{\{k-1\}}\boldsymbol{B}_{k-1} + \mathcal{N}(1, \boldsymbol{\Sigma}_{kk.\{k-1\}}),
\end{aligned}$$

with $\boldsymbol{\Gamma}_i = \boldsymbol{\Sigma}_{(i+1)(i+1).\{i\}}$, $i = 1, \ldots, k-1$. The distribution of $\boldsymbol{Y}$ is formed as the product of these distributions and there is a 1-1 correspondence between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{11}, (\boldsymbol{B}_i, \boldsymbol{\Gamma}_i)$, $i = 1, \ldots, k-1$.

We are now in a position to define the *GIW* distribution.

Definition 1 *The Generalized inverted Wishart distribution*
Suppose the q variables are partitioned into k sets as described above.Let $\boldsymbol{\Sigma}_{11}$ and pairs $(\boldsymbol{B}_i, \boldsymbol{\Gamma}_i)$, $i = 1, \ldots, k-1$ be mutually independent a priori. In addition suppose for matrices of constants $\boldsymbol{B}_{*i}, \boldsymbol{Q}_i, \boldsymbol{H}_i$, and scalars δ_i, $i = 1, \ldots, k-1$, as well as δ_0, Q_0,

$$\boldsymbol{\Sigma}_{11} \sim \mathcal{IW}(\delta_0, \boldsymbol{Q}_0) \tag{3}$$

$$\boldsymbol{B}_i \mid \boldsymbol{\Gamma}_i \sim \boldsymbol{B}_{*i} + \mathcal{N}(\boldsymbol{H}_i, \boldsymbol{\Gamma}_i) \tag{4}$$

$$\boldsymbol{\Gamma}_i \sim \mathcal{IW}(\delta_i + q_{\{i\}}, \boldsymbol{Q}_i). \tag{5}$$

Then $\boldsymbol{\Sigma}$ *is said to have a* $\mathcal{GIW}(\delta_0, \delta_i; \boldsymbol{B}_{*i}; \boldsymbol{H}_i, \boldsymbol{Q}_0, \boldsymbol{Q}_i,\ i = 1\ldots, k-1)$ *distribution,where* $\boldsymbol{\Sigma}$ *is defined through the recursive partition.*

The hyperparameters $(\delta_0, Q_0, \delta_i,\ \boldsymbol{B}_{*i}, \boldsymbol{H}_i, \boldsymbol{Q}_i,\ i = 1\ldots, k-1)$ involve a much larger number of distinct parameters than specified in an $\mathcal{IW}(\delta,\ \boldsymbol{G})$ for the $\boldsymbol{q} \times \boldsymbol{q}$ matrix $\boldsymbol{\Sigma}.$ Some of this excess is unneeded. If we retain the same mean structure as with the *IW* distribution then the additional span of the *GIW* family derives from two sources. The first is the varying shape parameters δ_i of its submodules,the second source of flexibility is found in the $\boldsymbol{H}_i,$ which determine the covariance structure for the $\boldsymbol{B}_i.$ Through the expectation of products of $\boldsymbol{Y}'$s from the submodules it may be seen that a suitable restriction of the hyperparameters is

$$\begin{aligned} \boldsymbol{Q}_0 &= \boldsymbol{G}_{11}, && (6)\\ \boldsymbol{Q}_i &= \boldsymbol{G}_{(i+1)(i+1).\{i\}}, && (7)\\ \boldsymbol{B}_{*i} &= \boldsymbol{G}^{-1}_{\{i\}\{i\}}\boldsymbol{G}_{\{i\}(i+1)}, && (8) \end{aligned}$$

for $i = 1\ldots, k-1$. This leaves $\boldsymbol{H}_i$ and δ_i free to specify involving an additional

$$\sum_{1}^{k-1} q_{\{i\}}(q_{\{i\}} - 1)/2 + k$$

parameters. For many purposes such flexibility may be superfluous and it will suffice to retain just the extra δ_i parameters with $\boldsymbol{H}_i$ for the ith regression module determined by the hyperparameters of earlier submodules in the sequence. Specifically we take

$$\boldsymbol{H}_i \ = \tau_{i-1}\boldsymbol{G}^{-1}_{\{i\}\{i\}},\ \ i = 1, \ldots, k-1, \tag{9}$$

with

$$\tau_i = \frac{\delta_i + q - 1}{\delta_{i+1} + q - 1},\ \ i = 0, 1, \ldots, k-2, \tag{10}$$

a slight change from Brown *et al* (1994a), motivated to preserve expected values and further explored below. This leaves the differing shape parameters $\delta_0, \delta_1, \ldots, \delta_{k-1}$. If these parameters were in turn taken to be equal to a common δ then we would be back to a standard $\mathcal{IW}(\delta; \boldsymbol{G})$ distribution. The independence of parameters in different submodules in the definition of the *GIW* above is indeed a feature of the *IW,* having its roots in the Bartlett decomposition of $\boldsymbol{\Sigma}.$

Garthwaite and Al-Awadhi (2001) give methods of assessing the the hyperparameters of the chain of submodules, from the first involving just a single inverted Wishart, then the sequence of $(k-1)$ regression modules are assessed involving parameters, $(\boldsymbol{B}_{*1}, \boldsymbol{H}_1;\delta_1,\boldsymbol{Q}_1)$ for the first of these up to $(\boldsymbol{B}_{*k-1}, \boldsymbol{H}_{k-1};\delta_{k-1},\boldsymbol{Q}_{k-1})$ for the last. They choose to allow flexibility in the distribution only via differing δ_i and specified the sequence of slope accuracies in terms of the parameters from the previous submodule,

$$\begin{aligned} \boldsymbol{H}_1 &= \tau_0\boldsymbol{Q}_0^{-1} && (11)\\ \boldsymbol{H}_{i+1} &= \tau_i\begin{pmatrix} \boldsymbol{H}_i^{-1} & \boldsymbol{H}_i^{-1}\boldsymbol{B}_{*i} \\ \boldsymbol{B}'_{*i}\boldsymbol{H}_i^{-1} & \boldsymbol{Q}_i + \boldsymbol{B}'_{*i}\boldsymbol{H}_i^{-1}\boldsymbol{B}_{*i} \end{pmatrix} && (12) \end{aligned}$$

for $i = 1\ldots, k-2$. This particular choice of the $\boldsymbol{H_i}$ involving the τ_i given by (10) was motivated by considering the form of prior to be what you would expect as a posterior from 'monotone' data on top of a vague Jeffreys invariant prior for $\boldsymbol{\Sigma}$. We explore this and other posterior forms in the next section.

2 Posterior distributions

We consider here four routes to a *GIW* posterior distribution:

(i) A *GIW* prior plus a full $(n \times q)$ data matrix $\boldsymbol{Y}$.

(ii) A vague prior plus a monotone data structure. This expresses the natural motivation of a *GIW* prior distribution.

(iii) An *IW* prior plus a monotone data structure.

(iv) A *GIW* prior plus a monotone data structure.

Since the submodules compose a succession of regressions with separate parameters all these cases can be arrived at straightforwardly by Bayes theorem. The structure is evident from the following lemma taken from Brown (1993) Appendix B.

Lemma 1 *Suppose* $\boldsymbol{Y}(n \times r)$ *and* $\boldsymbol{X}(n \times p)$ *are observed as realizationsof the model with given* $\boldsymbol{B}$,

$$\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B} \sim \mathcal{N}(\boldsymbol{I_n}, \boldsymbol{\Sigma}),$$

with $\boldsymbol{\Sigma}$ $(r \times r)$ *a temporarily assumed known covariance matrix. A priori assume given* $\boldsymbol{\Sigma}$ *that*

$$\boldsymbol{B} - \boldsymbol{B}_* \sim \mathcal{N}(\boldsymbol{H}, \boldsymbol{\Sigma}).$$

Then, given $\boldsymbol{\Sigma}, \boldsymbol{Y},$ *and* $\boldsymbol{X}$, *a posteriori*

$$\boldsymbol{B} - [\boldsymbol{W}\boldsymbol{B}_* + (\boldsymbol{I} - \boldsymbol{W})\hat{\boldsymbol{B}}] \sim \mathcal{N}((\boldsymbol{H}^{-1} + \boldsymbol{X}'\boldsymbol{X})^{-1}, \boldsymbol{\Sigma}) \tag{13}$$

where

$$\boldsymbol{W} = (\boldsymbol{H}^{-1} + \boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}^{-1}.$$

Thus the posterior mean of $\boldsymbol{B}$ *is a matrix weighted average of the prior mean* $\boldsymbol{B}_*$ *and the least-squares estimate*

$$\hat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}. \tag{14}$$

If in addition $\boldsymbol{\Sigma}$ *is unknown and a priori*

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\delta; \boldsymbol{Q})$$

then a posteriori,

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\delta + n;\ \boldsymbol{V} + \boldsymbol{Q} + (\hat{\boldsymbol{B}} - \boldsymbol{B}_*)'\{\boldsymbol{H} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\}^{-1}(\hat{\boldsymbol{B}} - \boldsymbol{B}_*) \tag{15}$$

where

$$\boldsymbol{V} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{B}}),$$

the least-squares residual error sum of products matrix.

2.1 Posterior for complete data

Suppose we have a *GIW* prior distribution and data $\boldsymbol{Y}(n \times q)$ whose rows are independently multivariate normal with mean zero and covariance matrix $\boldsymbol{\Sigma}$, namely

$$\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{I}_n, \boldsymbol{\Sigma}).$$

Then the posterior distribution is also $GIW(\tilde{\delta}_0, \tilde{\delta}_i;\, \tilde{\boldsymbol{B}}_{*i};\, \tilde{\boldsymbol{H}}_i, \tilde{\boldsymbol{Q}}_0, \tilde{\boldsymbol{Q}}_i,\ i = 1\ldots, k-1)$. Here $\tilde{\delta}_i = \delta_i + n,\ i = 0, 1\ldots, k-1$ and

$$\begin{array}{rclrcl}
\tilde{\boldsymbol{Q}}_0 &=& \boldsymbol{Q}_0 + \boldsymbol{Y}_1'\boldsymbol{Y}_1 & \tilde{\boldsymbol{H}}_i^{-1} &=& (\boldsymbol{H}_i^{-1} + \boldsymbol{Y}_{\{i\}}'\boldsymbol{Y}_{\{i\}}) \\
\tilde{\boldsymbol{Q}}_i &=& \boldsymbol{Q}_i + \boldsymbol{V}_i + [\hat{\boldsymbol{B}}_i - \boldsymbol{B}_{*i}]'\boldsymbol{A}_i[\hat{\boldsymbol{B}}_i - \boldsymbol{B}_{*i}] & \boldsymbol{A}_i &=& [\boldsymbol{H}_i + (\boldsymbol{Y}_{\{i\}}'\boldsymbol{Y}_{\{i\}})^{-1}]^{-1} \\
\boldsymbol{V}_i &=& (\boldsymbol{Y}_{i+1} - \boldsymbol{Y}_{\{i\}}\hat{\boldsymbol{B}}_i)'(\boldsymbol{Y}_{i+1} - \boldsymbol{Y}_{\{i\}}\hat{\boldsymbol{B}}_i) & \hat{\boldsymbol{B}}_i &=& (\boldsymbol{Y}_{\{i\}}'\boldsymbol{Y}_{\{i\}})^{-1}\boldsymbol{Y}_{\{i\}}'\boldsymbol{Y}_{i+1} \\
\boldsymbol{W}_i &=& (\boldsymbol{H}_i^{-1} + \boldsymbol{Y}_{\{i\}}'\boldsymbol{Y}_{\{i\}})^{-1}\boldsymbol{H}_i^{-1} & \tilde{\boldsymbol{B}}_{*i} &=& \boldsymbol{W}_i\boldsymbol{B}_{*i} + (\boldsymbol{I} - \boldsymbol{W}_i)\hat{\boldsymbol{B}}_i.
\end{array}$$

2.2 Monotone data driven prior distribution

It is natural to ask whether the *GIW* as a prior distribution can itself be thought of as a posterior distribution from a baseline prior such as the Jeffreys invariant prior distribution. For the generalized inverted Wishart the fundamental data structure is monotone. Such monotone data structures arise in spatial interpolation with air pollution data as in Le *et al* (1999) where monitoring stations for ozone levels are added to a network through time. Data is not available at all stations over the whole time period. This leads to what they call a staircase structure of data. We will summarise a result of Garthwaite and Al-Awadhi (2001) which shows that the *GIW* as a prior is itself a posterior for a vague prior with a staircase or monotone data structure. Suppose $q-$variate observations are independent normal with zero mean and covariance matrix $\boldsymbol{\Sigma}$, but for most observations only some (the first few) components of the response are known. Specifically, the first q_1 components of the first n_1 observations are known and their remaining components unknown; the first $q_1 + q_2 = q_{\{2\}}$ components of the next n_2 observations are known and their remaining components are unknown, $\ldots$, all $q_{\{k\}}$ components of the last n_k are known. Thus in the $n \times q$ matrix of data only the lower 'triangle' is observed, that is $(\boldsymbol{D}_{ij})$ for $i \leq j$ where $\boldsymbol{D}_{ij}$ is a $n_i \times q_j$ matrix of data.

Lemma 2 *Suppose the prior distribution for* $\boldsymbol{\Sigma}$ *is diffuse with probability function proportionalto* $\mid \boldsymbol{\Sigma}|^{-v/2}$ *and the data follow the staircase structure described above. Then the* *posterior distribution of* $\boldsymbol{\Sigma}$ *is GIW with,*

$$\boldsymbol{Q}_0 = \boldsymbol{S}_{11} \tag{16}$$

$$\delta_i = n_{i+1} + \dots n_k - v + 2,\ i = 0, 1, \dots, k-1, \tag{17}$$

and for $i = 1, \dots, k-1$,

$$\begin{aligned}
\boldsymbol{Q}_i &= \boldsymbol{S}_{(i+1)(i+1)} - \boldsymbol{S}_{(i+1)\{i\}}\boldsymbol{S}^{-1}_{\{i\}\{i\}}\boldsymbol{S}_{\{i\}(i+1)} \\
\boldsymbol{B}_{*i} &= \boldsymbol{S}^{-1}_{\{i\}\{i\}}\boldsymbol{S}_{\{i\}(i+1)} \\
\boldsymbol{H}_i &= \boldsymbol{S}^{-1}_{\{i\}\{i\}}
\end{aligned}$$

where $\boldsymbol{S}_{(i+1)\{i\}}$ *consists of the* $q_{i+1} \times q_{\{i\}}$ *cross products matrix of available* $n - n_{\{i\}}$ *observations. The other cross product matrices are similarly defined.*

The Jeffreys invariant prior has $v = q + 1$. This prior together with a monotone data structure were explored by Liu (1993), in the context of speeding up the EM algorithm for missing data problems. The posterior was characterised in this particular case (with the addition of a non-zero unknown mean).

Working backwards for a *GIW* distribution to correspond to a staircase data structure on a vague prior baseline then various constraints on the parameters would be needed. In particular, equation (17) implies that

$$\delta_0 \geq \delta_1 \geq \dots \geq \delta_{k-1}. \tag{18}$$

This suggests that a *GIW* distribution is likely to form a better representation of an expert's opinion if variables are ordered so that earlier components of $\boldsymbol{Y}$ are the ones the expert knows most about.

2.3 Posterior for monotone data structures

A monotone data structure can be combined with a *IW* prior to form a *GIW* posterior distribution. All that is required is to partition the *IW* conformally with the data partition. This dissection of the *IW* prior creates a *GIW* prior with hyperparameters in the submodules satisfying eqns (6), (7), (8), (9) together with equality of the prior δ's. The posterior then follows from that of each of the submodules. What's more a *GIW* prior with unequal ordered δ's satisfying (18) can clearly be combined with a monotone data structure for the same lexicographic order, by taking the partition of variables which is the intersection of the prior and data partitions. Bayes theorem is then applied to the submodules for updating to the posterior.

3 Working with the *GIW*

The *GIW* distribution of $\boldsymbol{\Sigma}$ is defined through the partition of variables and the regression submodules used in its construction. It is as straightforward to simulate from as is a standard *IW*. Moments of the distribution of $\boldsymbol{\Sigma}$ or functions of $\boldsymbol{\Sigma}$ may also often be calculated straightforwardly from the independence properties of parameters recursively defined, see Brown *et al* (1994a). Such expectations are a necessary ingredient of approaches the hyperparameter estimation based on the EM algorithm, as in Le *et al* (1999), an extension of its use for pollution interpolation with the IW prior in Brown *et al* (1994b). Predictive distributions are also readily available.

We have presented the GIW for sampling from a mean zero Gaussian. Extension to a non zero mean or even a linear model expectation (with covariates) is developed as the Gaussian- generalized inverted Wishart distribution in Le *et al* (1999).

References

Brown, P. J. (1993). *Measurement, Regression,and Calibration.* Oxford: Clarendon Press.

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

Brown, P. J., Le, N. D. and Zidek, J. V. (1994b). Multivariate spatial interpolation and exposure to air pollutants. *Canadian Journal of Statistics*, **22,** 489–509.

Connor, R. J. and Mosimann, J. E. (1969). Concepts of independence for proportions with a generalisation of the Dirichlet. *Journal of the American Statistical Association*, **64**.

Dawid, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, **68,** 265–274.

Garthwaite, P. H. and Al-Awadhi, S. A. (2001). Non-conjugate prior distribution assessment for multivarate normal sampling. *Journal Royal Statistical Society,* B, **63,** to appear.

Le, N. D., Sun, L. and Zidek, J. V. (1999). Bayesian spatial interpolation and backcasting using Gaussian- generalized inverted Wishart model. Technical Report 185. Statistics Dept., UBC Vancouver.

Liu, C. (1993). Bartlett's decomposition of the posterior distribution of the covariance for normal monotone ignorable missing data. *Journal of Multivariate Analysis*, **46,** 198–206.

Richard, J. F. and Steel, M. F. J. (1988). Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density. *Journal of Econometrics*, **38,** 7–37.