Stat Comput
DOI 10.1007/s11222-011-9295-3

Bayesian tests on components of the compound symmetry covariance matrix

Joris Mulder · Jean-Paul Fox

Received: 21 April 2011 / Accepted: 10 October 2011

Abstract Complex dependency structures are often conditionally modeled, where random effects parameters are used to specify the natural heterogeneity in the population. When interest is focused on the dependency structure, inferences can be made from a complex covariance matrix using a marginal modeling approach. In this marginal modeling framework, testing covariance parameters is not a boundary problem. Bayesian tests on covariance parameter(s) of the compound symmetry structure are proposed assuming multivariate normally distributed observations. Innovative proper prior distributions are introduced for the covariance components such that the positive definiteness of the (compound symmetry) covariance matrix is ensured. Furthermore, it is shown that the proposed priors on the covariance parameters lead to a balanced Bayes factor, in case of testing an inequality constrained hypothesis. As an illustration, the proposed Bayes factor is used for testing (non-)invariant intra-class correlations across different group types (public and Catholic schools), using the 1982 High School and Beyond survey data.

Keywords Bayes factor · Covariance matrices · Gibbs Sampler · Intra-class correlation · Compound symmetry · Savage-Dickey density ratio

This research was supported by grant NWO-VICI-453-05-002 of the Netherlands Organization for Scientific Research.

J. Mulder (✉)
Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands
e-mail: j.mulder3@uvt.nl

J.-P. Fox
Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, The Netherlands

1 Introduction

Data with a grouped structure, e.g., math scores of students from different schools or repeated measurements of different persons, are modeled such that the dependency structure is taken into account. In a conditional modeling approach, normally distributed observations within a group j are i.i.d. given the random group effect; that is,

$$y_{ij} = \mu + \mu_{0j} + e_{ij} \tag{1}$$

where

$$e_{ij} \sim \mathcal{N}(0, \sigma^2)$$
$$\mu_{0j} \sim \mathcal{N}(0, \tau).$$

Besides the common parameters, independent group-specific parameters are introduced with a common distribution function, usually normal, with a common mean and variance parameter. In this conditional modeling framework, the random effects parameters as well as the variance components need to be estimated to make inferences about the dependency structure.

The random intercept model in (1) is particularly popular in educational research to model the nesting of children in classes or in schools (e.g., Fox 2010; Gelman and Hill 2007; Raudenbush and Bryk 2002; Snijders and Bosker 1999). In such a two-level nesting, the within-class observations are positively correlated, which is represented by the intraclass correlation coefficient ($\tau/(\tau + \sigma^2)$). The random intercept model provides useful information through the partitioning of the total variation in between and within classes such as an estimate of the degree of dependence within each class.

In many cases the random effects are considered to be nuisance parameters and the structure of dependence is of main interest. When interest is focused on the dependency structure, inferences can be made from a complex covariance matrix using a marginal modeling approach. The implied marginal model defines the same dependency structure within a group (e.g., subject or school). Therefore, the observations are assumed to be multivariate normally distributed with a compound symmetry covariance matrix, that is,

$$\mathbf{y}_j \sim \mathcal{N}(\mu \mathbf{1}_{n_j}, \boldsymbol{\Sigma}_{n_j}), \quad \text{with } \boldsymbol{\Sigma}_{n_j} = \sigma^2 \mathbf{I}_{n_j} + \tau \mathbf{J}_{n_j}, \tag{2}$$

where μ is the overall mean across all groups, $\mathbf{I}_{n_j}$ is the $n_j \times n_j$ identity matrix, $\mathbf{J}_{n_j}$ is a $n_j \times n_j$ matrix containing ones. The vectors $\mathbf{y}_j$ are exchangeable for all J, which means that the vectors are i.i.d. given μ (e.g., sampling from an infinite population without replacement). Note that the covariance between two observations in the same class is τ under the random intercept model, (1), and the marginal model, (2). The marginal modeling approach is more general since complex dependency structures can be defined for which a conditional model may simply not exist, for example, when within-group observations are not homogenously correlated.

Testing a covariance structure using the marginal model (2) has certain advantages. For normally distributed observations, testing the existence of a random effect, i.e., $H_0 : \tau = 0$, is a boundary testing problem in the conditional model because $\tau \geq 0$. This is not the case in the marginal model because the covariance parameter τ can be negative in (2), which will be shown later on. Note that more general complex dependency structures can also be tested via the marginal model, where the corresponding conditional framework requires multiple random effects assumptions. For example, Verbeke et al. (2001) considered the study of longitudinal effects in the presence of cross-sectional effects where the longitudinal effects may be highly influenced by the cross-sectional effects. Then, the object is to identify longitudinal dependencies independent of any cross-sectional assumptions.

One problematic aspect is to specify prior distributions on the variance components τ and σ^2 such that positive definiteness of the entire covariance matrix is ensured. However, it will be shown that for a compound symmetry structure priors can be defined in such a way that an expression can be obtained for the posterior distribution of the covariance parameters. This allows one to test the support of a random effect, i.e., $H_0 : \tau = 0$ versus $H_1 : \tau > 0$ under the marginal modeling framework. If hypothesis H_1 is favored against H_0, the considered dependency structure is preferred and as a result the use of the conditional model is justifiable. For tests on τ under the conditional model, see for instance Kato and Hoijtink (2004) and Snijders and Bosker (1999). The test outcome can be informative for model building to justify any conditional independence assumptions.

A Bayesian test is described for testing invariance of τ across different group types, which is often implicitly assumed in the conditional modeling framework. In the case of A different types of groups, the test problem is considered of $H_0 : \tau_1 = \cdots = \tau_A$ against $H_1 : \tau_1 < \cdots < \tau_A$, where τ_a is the between-group variance of type $a = 1, \ldots, A$. Hypothesis H_0 states that the variance at the level of groups is equal across all group types and hypothesis H_1 states that there is a specific ordering. A numerical procedure is proposed to estimate the Bayes factor of H_0 against H_1 based on the prior and posterior sample of $\boldsymbol{\tau}$. The proposed methodology to calculate Bayes factors is based on the encompassing modeling approach (e.g., Berger and Mortera 1999; Klugkist and Hoijtink 2007). The procedure is illustrated using the 1982 High School and Beyond (HSB) survey data, which include information on math tests of 7,185 students nested within 90 public schools and 70 Catholic schools (Raudenbush and Bryk 2002). The Bayes factor is used to test whether students in Catholic schools are more alike than students in public schools with respect to their math performances.

The paper is organized as follows. First, prior densities are given for the covariance parameters that define a symmetric positive definite covariance matrix when dealing with a compound symmetry structure. The posterior density is derived for the covariance parameters, and a sampling mechanism is described for the marginal model parameters assuming a common nesting of groups. In a short example, covariance structures are marginally tested using two small data sets, which were discussed in Box and Tiao (1973). In the first data set, the random intercept model was applicable, and in the second this was not the case. In the proposed modeling framework this is tested by evaluating the marginal posterior probability of the null hypothesis $\tau > 0$. In the following section, a generalization is made for testing heterogenous compound symmetry structures where equality or inequality constrained hypotheses of covariance parameters are considered. It is discussed how to test homogeneity of τ across different group types using the Bayes factor as a selection criterion. The test is also applied to the HSB data to test homogeneity of τ across public and Catholic schools.

2 Testing compound symmetry: prior specification

A multivariate normal distribution with compound symmetry covariance matrix can be used for modeling a vector with dependent continuous observations with approximately equal variances and covariances. When jointly modeling J vectors of length n_j, independently and identically distributed according to (2), the marginal model can be written as

$$\mathbf{y} = N(\mu \mathbf{1}_N, \boldsymbol{\Sigma}_N), \quad \text{with}$$

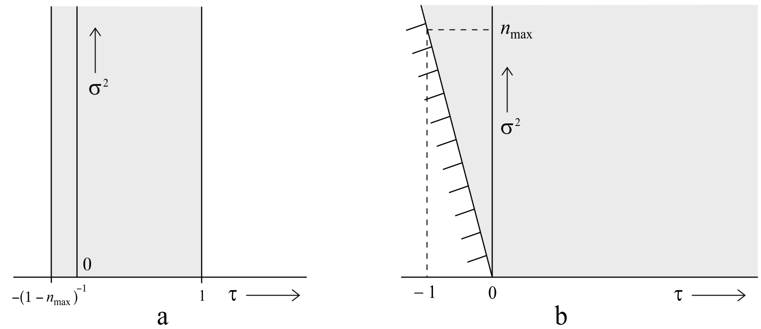

Fig. 1 In (**a**) the allowed parameter space of (τ, σ^2) for the compound symmetry parameterization $\boldsymbol{\Sigma}_{n_j}^{CS1} = \sigma^2((1-\tau)\mathbf{I}_{n_j} + \tau\mathbf{J}_{n_j})$, and in (**b**) for the parameterization $\boldsymbol{\Sigma}_{n_j}^{CS2} = \sigma^2\mathbf{I}_{n_j} + \tau\mathbf{J}_{n_j}$

$$\boldsymbol{\Sigma}_N = \begin{bmatrix} \boldsymbol{\Sigma}_{n_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Sigma}_{n_J} \end{bmatrix} \tag{3}$$

where $\mathbf{y}' = (\mathbf{y}'_1, \ldots, \mathbf{y}'_J)$, μ is the overall mean across observations, $\boldsymbol{\Sigma}_{n_j}$ is a $n_j \times n_j$ compound symmetry covariance matrix, and $N = \sum_j n_j$.

The compound symmetry structure implies identical diagonal elements and identical off-diagonal elements. Two commonly used parameterizations of the compound symmetry structure are used,

$$\boldsymbol{\Sigma}_{n_j}^{CS1} = \sigma^2 \begin{bmatrix} 1 & \tau & \cdots & \tau \\ \tau & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tau \\ \tau & \cdots & \tau & 1 \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{\Sigma}_{n_j}^{CS2} = \begin{bmatrix} \sigma^2+\tau & \tau & \cdots & \tau \\ \tau & \sigma^2+\tau & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tau \\ \tau & \cdots & \tau & \sigma^2+\tau \end{bmatrix}, \tag{4}$$

which can be written as

$$\boldsymbol{\Sigma}_{n_j}^{CS1} = \sigma^2((1-\tau)\mathbf{I}_{n_j} + \tau\mathbf{J}_{n_j}),$$
$$\boldsymbol{\Sigma}_{n_j}^{CS2} = \sigma^2\mathbf{I}_{n_j} + \tau\mathbf{J}_{n_j},$$

respectively, where $\mathbf{I}_{n_j}$ is the identity matrix of dimension $n_j \times n_j$ and $\mathbf{J}_{n_j}$ is a matrix of dimension $n_j \times n_j$ containing ones. Note that the multivariate model in (3) can be written as in (2) with either $\boldsymbol{\Sigma}_{n_j}^{CS1}$ or $\boldsymbol{\Sigma}_{n_j}^{CS2}$ as the covariance matrix.

In the parametrization $\boldsymbol{\Sigma}_{n_j}^{CS2}$, the intra-class correlation is parameterized as $\tau/(\tau+\sigma^2)$ and the total variance as $\sigma^2+\tau$. This intra-class correlation coefficient is represented by τ and the total variance by σ^2 in the parametrization of $\boldsymbol{\Sigma}_{n_j}^{CS1}$. Therefore, in this parameterization τ will also be referred to as the intra-class correlation.

In a Bayesian modeling approach, the likelihood model in (3) requires a specification of the prior for the variance components σ^2 and τ that results in a positive-definite covariance matrix $\boldsymbol{\Sigma}_N$. The covariance matrix is positive definite if the blocks on the diagonal, i.e., $\boldsymbol{\Sigma}_{n_j}$ for $j = 1, \ldots, j$, are positive definite. For the parameterization $\boldsymbol{\Sigma}_{n_j}^{CS1}$, this is the case for the set $\{\tau, \sigma^2 | -(n_{\max}-1)^{-1} < \tau < 1, \sigma^2 > 0\}$, with $n_{\max} = \max_j(n_j)$. For the parameterization $\boldsymbol{\Sigma}_{n_j}^{CS2}$, this is the case for $\{\tau, \sigma^2 | \tau > -\sigma^2/n_{\max}, \sigma^2 > 0$. Figure 1 gives graphical representations of the allowed parameter spaces.

For the parameterization $\boldsymbol{\Sigma}_{n_j}^{CS2}$, a noninformative improper prior of the variance components $\pi(\sigma^2, \tau) = \pi(\sigma^2)\pi(\tau|\sigma^2) \propto \sigma^{-2}(\tau + \sigma^2/n_{\max})^{-1}$ was used by Box and Tiao (1973). This prior cannot be used for Bayes factor testing because it is improper (Jeffreys 1961). Different complex methods have been proposed for constructing automatic proper data-based priors, which are relatively noninformative and located around the likelihood that lead to so-called default Bayes factors (e.g., O'Hagen 1995; Berger and Pericchi 1996).

The specification of a prior for the variance components that results a positive definite covariance matrix with compound symmetry structure is simpler when using the parameterization $\boldsymbol{\Sigma}_{n_j}^{CS1}$. For this parameterization we consider a noninformative improper prior for estimation and a noninformative proper prior for hypothesis testing using the Bayes factor. The noninformative improper prior that is used is given by

$$\pi(\mu, \tau, \sigma^2) \propto \sigma^{-3}(1 + (n_{\max}-1)\tau)^{-1}(1-\tau)^{-1}. \tag{5}$$

This prior corresponds with the reference prior where the parameter τ is considered as more important than the parameters (μ, σ^2) (Berger and Bernardo 1992; Chung and Dey 1998). This prior is optimal in the sense that inference is maximally based on the data at hand.

For testing constrained hypotheses on intra-class correlation coefficients, which is discussed in Sect. 3, we consider the following proper prior

$$\pi(\mu, \tau, \sigma^2) = \pi(\mu)\pi(\tau)\pi(\sigma^2) \tag{6}$$

with

$$\pi(\mu) = U(-1\text{e}10, 1\text{e}10)$$

$\pi(\tau) = U(-(n_{\max} - 1)^{-1}, 1)$

$\pi(\sigma^2) = \text{inv-}\chi^2(1, 1).$

As will be shown, this prior results in Bayes factors that are balanced for testing inequality constrained hypotheses on multiple τ's. This prior can be easily generalized to a hierarchical prior such that the prior parameters are modeled at a higher level. The hyperprior for the prior parameters can be used to specify the parameter region instead of fixing the prior parameters.

Another useful property of the parameterization $\boldsymbol{\Sigma}_{n_j}^{CS1}$ is that τ corresponds with the intra-class correlation, which is an important quantity for grouped data. A large intra-class correlation implies a strong resembles of the observations within a group, and therefore corresponds with a relatively small within-group variance and a relatively large between-group variance. Furthermore, σ^2 can be interpreted as the total variance, i.e., the sum of the within-group variance and the between-group variance. Finally note that the conditional version of the marginal model with parameterization $\boldsymbol{\Sigma}_{n_j}^{CS1}$ would be $y_{ij} = \mu + \mu_{0j} + \epsilon_{ij}$, where $\epsilon_{ij} \sim N(0, \sigma^2(1-\tau))$ and $\mu_{0j} \sim N(0, \sigma^2\tau)$.

2.1 Posterior computation: sampling of the mean and variance components

Let the compound symmetry covariance matrix $\boldsymbol{\Sigma}_{n_j}^{CS1}$ be denoted by $\sigma^2\boldsymbol{\Sigma}_{\tau,n_j}$, with $\boldsymbol{\Sigma}_{\tau,n_j} = (1-\tau)\mathbf{I}_{n_j} + \tau\mathbf{J}_{n_j}$, and $\boldsymbol{\Sigma}_{\tau,N} = \text{diag}(\boldsymbol{\Sigma}_{\tau,n_1}, \ldots, \boldsymbol{\Sigma}_{\tau,n_J})$ where $N = \sum_j n_j$.

The likelihood function of the marginal model in (3) is given by

$$
\begin{aligned}
&f(\mathbf{y} \mid \tau, \sigma^2, \mu) \\
&\quad = (2\pi\sigma^2)^{-N/2} |\boldsymbol{\Sigma}_{\tau,N}|^{-1/2} \\
&\qquad \times \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mu)' \boldsymbol{\Sigma}_{\tau,N}^{-1} (\mathbf{y} - \mu) \right\},
\end{aligned} \tag{7}
$$

where $\boldsymbol{\Sigma}_{\tau,N}^{-1} = \text{diag}(\boldsymbol{\Sigma}_{\tau,n_1}^{-1}, \ldots, \boldsymbol{\Sigma}_{\tau,n_J}^{-1})$ and $\boldsymbol{\Sigma}_{\tau,n_J}^{-1} = (1-\tau)^{-1}\mathbf{I}_{n_j} - \frac{\tau}{(1-\tau)(1+(n_j-1)\tau)}\mathbf{J}_{n_j}$.

2.1.1 *Multiple-blocked MCMC scheme*

A straightforward Markov chain Monte Carlo (MCMC) implementation is based on the full conditional posterior distribution of the model parameters and iteratively draws are made from the full conditionals. Therefore, the conditional posterior distributions of μ, σ^2 and τ are derived using the reference prior in (5) and the likelihood in (7).

The conditional posterior of μ is normally distributed given by $\pi(\mu \mid \tau, \sigma^2, \mathbf{y}) = N((\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N)^{-1} \mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{y}, (\mathbf{1}_N' \boldsymbol{\Sigma}_N^{-1} \mathbf{1}_N)^{-1})$. For equal group-sizes, $n_j = n, \forall j$, the conditional posterior density simplifies to

$$
\begin{aligned}
&\pi(\mu \mid \tau, \sigma^2, \mathbf{y}) \\
&\quad = N\left(\bar{y}, \sigma^2 \left(\frac{Jn}{1-\tau} - \frac{Jn^2\tau}{(1-\tau)((1-\tau)+n\tau)} \right)^{-1} \right),
\end{aligned}
$$

where $\bar{y}$ is the overall mean of $\mathbf{y}$.

The conditional posterior of σ^2 has a scaled-inverse χ^2 distribution. First note that

$$
\begin{aligned}
&\mathbf{y}_j \sim N\left(\mu\mathbf{1}_{n_j}, \sigma^2(\tau\mathbf{J}_{n_j} + (1-\tau)\mathbf{I}_{n_j})\right) \\
&\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{B}'\mathbf{y}_j \sim N\left(\mu\boldsymbol{\Lambda}^{-\frac{1}{2}}\mathbf{B}'\mathbf{1}_{n_j}, \sigma^2\mathbf{I}_{n_j}\right)
\end{aligned} \tag{8}
$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues of $\tau\mathbf{J}_{n_j} + (1-\tau)\mathbf{I}_{n_j}$, given by $1 + (n_j - 1)\tau$ (with multiplicity 1) and $1-\tau$ (with multiplicity $n_j - 1$), and $\mathbf{B}$ is a $n_j \times n_j$ matrix with the orthogonal eigenvectors as columns. Consequently, by multiplying the likelihood with the noninformative improper prior in (5), the conditional posterior yields

$$
\pi(\sigma^2 \mid \tau, \mu, \mathbf{y}) = \text{inv-}\chi^2\left(s_{\sigma^2}^2, N+1\right), \tag{9}
$$

where

$$
s_{\sigma^2}^2 = (N+1)^{-1} \sum_j (\mathbf{y}_j' - \mu\mathbf{1}_{n_j}')\mathbf{B}\boldsymbol{\Lambda}^{-1}\mathbf{B}'(\mathbf{y}_j - \mu\mathbf{1}_{n_j}).
$$

The conditional posterior of τ does not have a common distribution, and therefore, τ needs to be sampled using a Metropolis-Hastings (M-H) step (Metropolis et al. 1953; Hastings 1970). The kernel of the conditional distribution of τ is proportional to the likelihood in (7). Browne (2006) explored how to sample variance components using different Metropolis-Hastings algorithms. In the illustration discussed below, a truncated normal proposal density in the interval $(-(n_{\max} - 1)^{-1}, 1)$ with standard error of 0.35 was used resulting in an acceptance rate of approximately 0.47. Note that due to the nonsymmetric proposal density the acceptance bound in the Metropolis-Hastings sampling step also depends on this proposal density.

2.1.2 *Single-blocked MCMC scheme*

It is often better to block highly correlated parameters such that, for example, proposal densities can be defined that improve the mixing and convergence properties of the algorithm. Following, among others, Andrieu and Thomas (2008) and Chib and Greenberg (1995), an adaptive M-H algorithm is defined such that a multivariate proposal density is optimized during a finite period of the chain. This leads to a more efficient M-H algorithm, when the target distribution is faster explored due to better proposals. This is not always the case and the behavior of the chains should be guided using MCMC diagnostic methods (Roberts and Sahu 1997).

The implementation of a single-block adaptive M-H algorithm requires a proposal density for μ, τ, and σ^2. As discussed in, for instance, Chib and Greenberg (1998), the proposal density should be tailored to the target density which in our case is $p(\mu,\tau,\sigma^2|\mathbf{y}) \propto \pi(\mu,\tau,\sigma^2) f(\mathbf{y}|\mu,\tau,\sigma^2) I_{\{(-(n_j-1)^{-1},1)\}}(\tau) I_{\{(0,\infty)\}}(\sigma^2)$. This can be done using the following approximation of the target density.

First, in each group, $\tilde{y}_{j1}$ is computed as the sum of the first half w_j observations, where $w_j = n_j/2$ if n_j is even and $w_j = n_j/2 + 0.5$ if n_j is odd. The sum of the other $n_j - w_j$ observations is denoted as $\tilde{y}_{j2}$. Second, the bivariate normal distribution of $\tilde{y}_{j1} \mid \tilde{y}_{j2}$ is used to approximate the likelihood of the model parameters, which is given by

$$\tilde{\mathbf{y}}_1 \mid \tilde{\mathbf{y}}_2 = N\left(\tilde{\mathbf{X}}\begin{bmatrix}\mu\\ \tau\end{bmatrix}, \sigma^2\tilde{\mathbf{D}}\right), \tag{10}$$

where the jth row of $\tilde{\mathbf{X}}$ is (w_j, z_{j1}) and $\tilde{\mathbf{D}}$ is a diagonal matrix where the (j,j)th element equals z_{j2}. Furthermore,

$$\begin{aligned} z_{j1} &= w_j(n_j - w_j)(\tilde{y}_{j2} - \mu(n_j - w_j)) \\ &\quad \times ((n_j - w_j)(n_j - w_j - 1)\tau + n_j - w_j)^{-1} \\ z_{j2} &= w_j(w_j - 1)\tau + w_j - \tau^2 w_j^2 (n_j - w_j)^2 \\ &\quad \times ((n_j - w_j)(n_j - w_j - 1)\tau + n_j - w_j)^{-1}. \end{aligned}$$

From (10), independent proposal densities for $(\mu,\tau)'$ and σ^2, can be derived. It follows that,

$$\begin{aligned} q(\mu^*, \tau^*) &= N(\mathbf{m}_1, \rho_1^2 \mathbf{V}) \\ q(\sigma^{2*}) &= N(m_2, \rho_2^2 v), \end{aligned}$$

where $\mathbf{m}_1$ is the posterior mean and $\mathbf{V}$ is the posterior covariance matrix. They can be obtained from (10) by substituting $\sigma^2 = \sigma^{2(s)}$, where s denotes the MCMC iteration number. Furthermore, m_2 is the posterior mean and v is the posterior variance of the inverse gamma distribution of σ^2. They can be obtained from (10) by substituting $(\mu,\tau) = (\mu^{(s)}, \tau^{(s)})$ multiplied with the improper prior σ^{-2}.

The proposal density needs to be tuned during a finite-period of the chain using the tuning parameters ρ_1 and ρ_2. This tuning is carried out until a beforehand chosen acceptance rate is established. In iteration $s+1$, the proposal $\mu^*, \tau^*, \sigma^{2*}$ is accepted when

$$u \le \frac{\frac{p(\mu^*,\tau^*,\sigma^{2*}|\mathbf{y})}{q(\mu^*,\tau^*,\sigma^{2*}|\mu^{(s)},\tau^{(s)},\sigma^{2(s)})}}{\frac{p(\mu^{(s)},\tau^{(s)},\sigma^{2(s)}|\mathbf{y})}{q(\mu^{(s)},\tau^{(s)},\sigma^{2(s)}|\mu^*,\tau^*,\sigma^{2*})}}, \tag{11}$$

where u is an observation from random variable U, which is uniformly distributed in the interval [0, 1]. Note that the proposal density q do not have to be truncated in the region $\{\tau, \sigma^2\} = (-(n_j-1)^{-1}, 1) \times (0,\infty)$ because these bounds are incorporated in the target density p. For this reason, the proposals are accepted with probability zero when they fall outside the allowed region.

2.2 Illustration

As an illustration we consider two data sets from Box and Tiao (1973, pp. 246–247), which contained $J = 6$ groups with $n_j = 5$ observations per group. These data sets have been analyzed more often, for instance, by Gelfand et al. (1990), who used a random intercept model to fit the data. The purpose of the illustration is to justify our prior choice, to show that the multiple-blocked M-H sampler and the one-blocked sampler (Sect. 2.1) work effectively to fit the data, and to show how to test the hypothesis $H_0 : \tau \le 0$ against $H_1 : \tau > 0$. Note here that a random effects model is only appropriate under H_1. Hypothesis H_0 will be tested by evaluating the posterior probability

$$\begin{aligned} P(H_0 \mid \mathbf{y}) &= P(\tau \le 0 \mid \mathbf{y}) \\ &= \int_{\tau \le 0} \pi(\tau \mid \mathbf{y}) d\tau \\ &\approx S^{-1} \sum_s I(\tau^{(s)} \le 0), \end{aligned} \tag{12}$$

where $\tau^{(s)}$ is the sth posterior draw of τ under the marginal model (3). The null hypothesis is accepted when the posterior probability of the null is greater than the posterior probability of the alternative.

A posterior sample of size 100,000 was obtained using the multiple-blocked M-H sampler and the one-blocked sampler discussed in the previous section. In the multiple blocked sampler, τ was sampled using a normal proposal density centered at the previous draw $\tau^{(s)}$ with a standard deviation of 0.35 resulting in an acceptance rate of approximately 0.47 for both data sets. The tuning parameters in the one-blocked sampler were chosen equal to $\rho_1 = 0.9$ and $\rho_2 = 0.6$ for both data sets resulting in acceptance rates of approximately 0.22 and 0.18 for data sets 1 and 2, respectively. In the lower panels of Fig. 2, the resulting posterior sample densities of τ are presented using the multiple-blocked sampler (solid line) and the one-blocked sampler (dashed line). The figure illustrates that the posterior samples are equivalent.

In the upper panels of Fig. 2, contour plots are given based on the posterior sample of (τ, σ^2) based on the multiple-blocked sampler (the one-blocked sampler was equivalent). The plots show that the posterior is well-behaved and is not multi-modal when using the reference prior in (5). This can be seen as an important justification of our prior choice.

Finally, the hypothesis $H_0 : \tau \le 0$ was tested against $H_1 : \tau > 0$ based on the posterior draws of τ. The posterior probability estimates $P(H_0 \mid \mathbf{y})$ were equal to 0.005 and

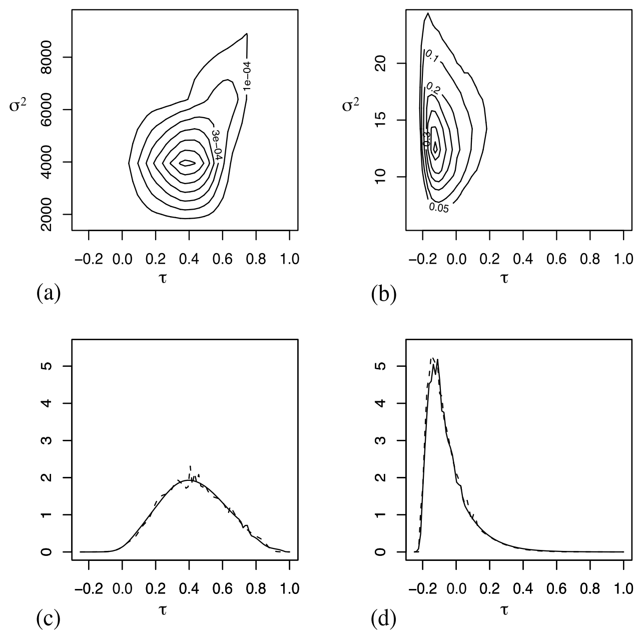

Fig. 2 Upper panels: Posterior contour plot of (τ, σ^2) based on data set 1 (*left*) and data set 2 (*right*) using the multiple-blocked M-H sampler. Lower panels: Posterior sample densities of τ for the data set 1 (*left*) and data set 2 (*right*) using the multiple-blocked M-H sampler (*solid line*) and the one-blocked M-H sampler (*dashed line*)

0.747 for data sets 1 and 2, respectively. Hence, hypothesis H_0 was rejected for data set 1 and accepted for data set 2. Based on these results it was concluded that the use of a random intercept model is justifiable for data set 1 but not for data set 2.

3 Testing invariance of the compound symmetry structure

Homogeneity of the intra-class correlation parameter, τ, is generally assumed when using the random effects model (1). However, there are numerous examples where the assumption of homogeneity of the variance components across groups is violated (Votaw 1948; Szatkowski 1982). The case will be considered that the observations are nested within groups, a compound symmetry structure holds, but the strength of within-group correlation can vary across groups. This violates the invariance assumption of common intra-class correlations across groups.

As a motivating example, the 1982 High School and Beyond (HSB) survey data are considered, which include information of mathematical grades of 7,185 students nested within 90 public schools and 70 Catholic schools (Raudenbush and Bryk 2002). The general assumption of an invariant compound structure across schools will be investigated. For example, it will be tested whether the grades of students within Catholic schools are more alike than the grades of students within public schools. In Sect. 4, the example is described in more detail.

Assume that the observations $\mathbf{y}_j$ are multivariate normally distributed, where the covariance matrix has a compound symmetry structure with a sector-specific covariance parameter τ_a $(a = 1, \ldots, A)$; that is,

$$\mathbf{y}_j \sim N\big(\mu, \sigma^2((1-\tau_{a_j})\mathbf{I}_{n_j} + \tau_{a_j}\mathbf{J}_{n_j})\big), \tag{13}$$

where $a_j = a$ if group j corresponds to sector a. Then, the following hypotheses can be formulated

$$\begin{aligned} &H_0 : \tau_1 = \cdots = \tau_A \\ &H_1 : \tau_1 < \cdots < \tau_A \\ &H_u : \tau_1, \ldots, \tau_A. \end{aligned} \tag{14}$$

Under the equality constrained hypothesis H_0, it is assumed that the compound symmetry structure is invariant over sectors. Hence, homogenous intra-class correlations are assumed under H_0. Under the inequality constrained hypothesis H_1, an ordering is defined on the sector-specific intra-class correlations. Under the unconstrained hypothesis H_u, sector-specific intra-class correlations are assumed.

The Bayes factor of hypothesis H_t against hypothesis $H_{t'}$ is defined as the ratio of the marginal likelihoods under the two hypotheses, i.e.,

$$B_{tt'} = \frac{m_t(\mathbf{y})}{m_{t'}(\mathbf{y})}, \tag{15}$$

with

$$m_t(\mathbf{y}) = \int f_t(\mathbf{y} \mid \boldsymbol{\theta}_t)\pi_t(\boldsymbol{\theta}_t)d\boldsymbol{\theta}_t,$$

where $f_t(\mathbf{y} \mid \boldsymbol{\theta}_t)$ is the likelihood of the data under H_t given its parameters $\boldsymbol{\theta}_t$. The Bayes factor $B_{tt'}$ can be interpreted as the amount of evidence in favor of hypothesis H_t against $H_{t'}$.

Hence, in order to obtain the Bayes factors between these hypotheses, the marginal likelihoods need to be calculated. Because this calculation can be computationally intensive, different methods have been proposed for computing marginal likelihoods, e.g., Chib and Jeliazkov (2001), Kass and Raftery (1995), and the references therein.

As was shown by Klugkist and Hoijtink (2007), however, the Bayes factors between the hypotheses H_0, H_1, and H_u can be obtained without computing the marginal likelihoods because the constrained hypotheses H_0 and H_1 are nested within the larger, unconstrained hypotheses H_u. It will be shown that analytical expressions for the Bayes factor concerning inequality (Sect. 3.1) and equality (Sect. 3.2) constrained compound symmetry structures can be derived. This will greatly simplify the computation of the Bayes factor. However, an adjustment is needed for the standard Bayes factor when testing inequality constrained hypotheses. For the general case, a numerical procedure will be proposed (Sect. 3.2.1) based on a posterior (and prior) sample of the parameter vector of interest.

3.1 Balanced BF for inequality constrained compound symmetry structures

It is generally stated that the Bayes factor works as an Ockham's razor, which means that it balances fit and complexity when quantifying the relative degree of support of two hypotheses. However, the choice of the prior is essential in order for this to hold. This is for instance illustrated by Bartlett's paradox for testing equality constrained hypotheses (Bartlett 1957; Jeffreys 1961) or by Mulder et al. (2010) for testing inequality constrained hypothesis. Berger and Pericchi (1996) use the term 'balanced' for Bayes factors that effectively balances fit and complexity. When the interest is in testing simple order restrictions of the parameters of interest, such as $H_1 : \tau_1 < \cdots < \tau_A$ in (14), we shall adopt the following definition for a balanced Bayes factor.

Definition 1 A Bayes factor is called balanced for testing simple order restrictions of A parameters when the prior probability that each ordering holds is equal across all $A!$ possible permutations.

Our goal is to obtain a Bayes factor that is balanced for testing order constrained hypotheses on $\boldsymbol{\tau}$. In this section, it will be discussed how to specify a prior that satisfies this requirement.

A straightforward choice for the prior of $(\boldsymbol{\tau}, \sigma^2, \mu)$ is to use a uniform prior for $\boldsymbol{\tau}$, which is proper and noninformative, and vague proper priors for σ^2 and μ, i.e.,

$$\pi_u(\tau_1, \ldots, \tau_A, \sigma^2, \mu) = \prod_a \pi_u(\tau_a)\pi_u(\sigma^2)\pi_u(\mu) \tag{16}$$

with

$$\pi_u(\tau_a) = U(-(n_{\max,a} - 1)^{-1}, 1), \quad \text{for } a = 1, \ldots, A$$

$$\pi_u(\sigma^2) = \text{inv-}\chi^2(1, 1)$$

$$\pi_u(\mu) = U(-1\text{e}10, 1\text{e}10),$$

where $n_{\max,a}$ is the size of the largest group in sector a, the prior of σ^2 is a scaled-inverse χ^2 distribution, and a vague uniform distribution is chosen for μ (Sect. 2).

Under the inequality constrained hypothesis H_t, the prior is a truncated version of the prior under H_u, i.e., $\pi_t(\boldsymbol{\tau}, \sigma^2, \mu) = c_t^{-1}\pi_u(\boldsymbol{\tau}, \sigma^2, \mu)I(\boldsymbol{\tau} \in H_t)$ (Berger and Mortera 1999; Klugkist and Hoijtink 2007). In this notation, c_t is a normalizing constant, which is equal to the prior probability that the inequality constraints of H_t hold under the unconstrained hypothesis H_u. The normalizing constant can be expressed as $c_t = \int_{\boldsymbol{\tau} \in H_t} \pi_u(\boldsymbol{\tau})\partial\boldsymbol{\tau}$. This prior specification results in a posterior density under H_t that is also a truncated version of the posterior density under H_u, i.e., $\pi_t(\tau_1, \ldots, \sigma^2, \mu|\mathbf{y}) = f_t^{-1}\pi_u(\tau_1, \ldots, \sigma^2, \mu|\mathbf{y})I(\boldsymbol{\tau} \in H_t)$. The normalizing constant $f_t = \int_{\boldsymbol{\tau} \in H_t} \pi_u(\boldsymbol{\tau}|\mathbf{y})\partial\boldsymbol{\tau}$ equals the posterior probability that the inequality constraints of H_t under the larger unconstrained hypothesis H_u.

For this prior specification, Klugkist and Hoijtink (2007) showed that the Bayes factor of the inequality constrained hypothesis H_t against the unconstrained model H_u can be expressed as the posterior probability that the inequality constraints of H_t hold divided by the prior probability that the inequality constraints of H_t hold, i.e.,

$$B_{tu} = \frac{f_t}{c_t} = \frac{\int_{\boldsymbol{\tau} \in H_t} \pi_u(\boldsymbol{\tau} \mid \mathbf{y})\partial\boldsymbol{\tau}}{\int_{\boldsymbol{\tau} \in H_t} \pi_u(\boldsymbol{\tau})\partial\boldsymbol{\tau}}, \tag{17}$$

where the numerator is the posterior probability that the inequality constraints of H_t hold, which is as a measure of *relative fit* of the inequality constrained hypothesis H_t (in comparison to H_u), and the denominator is the prior probability that the inequality constraints of H_t hold, which is as a measure of *relative complexity* of H_t (in comparison to H_u). When the same unconstrained prior π_u is used for calculating B_{tu} and $B_{t'u}$ for two inequality constrained hypotheses H_t and $H_{t'}$, it holds that $B_{tt'} = B_{tu}/B_{t'u}$.

Proposition 1 *A uniform prior on $\boldsymbol{\tau}$ under the unconstrained hypothesis H_u does not result in a balanced Bayes factor for testing order constraints of $\boldsymbol{\tau}$ when $n_{\max,a_1} \neq n_{\max,a_2}$ for at least one pair $a_1 \neq a_2$.*

Proof Let $\mathcal{Q}$ be the set of all $A!$ disjoint subspaces implied by simple order constraints on $\boldsymbol{\tau}$, e.g., $Q_t = \{\boldsymbol{\tau}|\tau_1 < \cdots < \tau_A\} \in \mathcal{Q}$. Hence, $\bigcup_{t=1}^{A!} Q_t = (-\frac{1}{n_{\max,1}}, 1) \times \cdots \times (-\frac{1}{n_{\max,A}}, 1)$ is the complete unconstrained space of $\boldsymbol{\tau}$. In order for the Bayes factor to be balanced, $P_{\pi_u}(\boldsymbol{\tau} \in Q_t)$ must be equal for all $t = 1, \ldots, A!$. Without losing generality let us assume that $n_{\max,1} < n_{\max,2}$. Consequently, $P_{\pi_u}(\tau_1 < \tau_2 < \tau_3 < \cdots < \tau_A) > P_{\pi_u}(\tau_2 < \tau_1 < \tau_3 < \cdots < \tau_A)$, and therefore, the Bayes factor will not be balanced for testing simple order constraints of $\boldsymbol{\tau}$ (see Fig. 3 for a sketch for $A = 2$). □

Two modifications can be considered to obtain a balanced Bayes factor for testing order constraints of $\boldsymbol{\tau}$ in a marginal modeling framework. First, the uniform prior for $\boldsymbol{\tau}$ under H_u can be changed such that $P_{\pi_u}(\boldsymbol{\tau} \in Q_{t_2}) = P_{\pi_u}(\boldsymbol{\tau} \in Q_{t_1})$ $\forall t_1, t_2 = 1, \ldots, A!$. Second, the prior probability of each ordering of the elements of $\boldsymbol{\tau}$ must be calculated by adding $\boldsymbol{\tau} > \mathbf{0}$ as extra constraints, i.e., $P_{\pi_u}(\boldsymbol{\tau} \in Q_{t_1} \cap \boldsymbol{\tau} > \mathbf{0}) = P_{\pi_u}(\boldsymbol{\tau} \in Q_{t_2} \cap \boldsymbol{\tau} > \mathbf{0})$ must be equal $\forall t_1, t_2 = 1, \ldots, A!$. We adopt the second method to avoid (unnecessary) complex

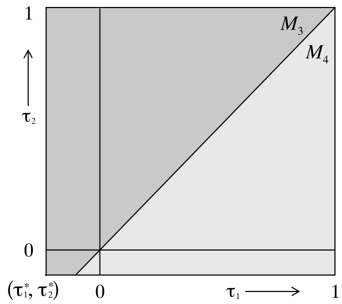

Fig. 3 Allowed parameter space of $\mathcal{M}_3 : \tau_1 < \tau_2$ and $\mathcal{M}_4 : \tau_1 > \tau_2$ with $(\tau_1^*, \tau_2^*) = (-\frac{1}{n_{\max,1}-1}, -\frac{1}{n_{\max,2}-1})$. If $n_{\max,1} < n_{\max,2}$ and assuming a uniform distribution over $(-\frac{1}{n_{\max,1}-1}, 1) \times (-\frac{1}{n_{\max,2}-1}, 1)$, the probability mass in the constrained spaces of $\mathcal{M}_3$ is larger than the probability mass in the space of $\mathcal{M}_4$, resulting in a biased Bayes factor towards model $\mathcal{M}_4$

prior specification for $\pi_u(\boldsymbol{\tau})$. Furthermore, the restriction $\boldsymbol{\tau} > \mathbf{0}$ is reasonable because ordering intraclass correlations only makes sense if the $\tau_a > 0, \forall a = 1, \ldots, A$.

Proposition 2 *A uniform prior on $\boldsymbol{\tau}$ under the unconstrained hypothesis H_u results in a balanced Bayes factor for testing order constraints of $\boldsymbol{\tau}$ by adding $\boldsymbol{\tau} > \mathbf{0}$ as extra set of constraints.*

Proof The prior probability of a simple ordering of the A τ's and $\boldsymbol{\tau} > \mathbf{0}$ under a uniform prior under H_u equals

$$P_{\pi_u}(\boldsymbol{\tau} \in Q_t \cap \boldsymbol{\tau} > \mathbf{0}) = \left(A! \prod_{a=1}^{A} \frac{n_{\max,a}}{n_{\max,a} - 1} \right)^{-1},$$

for all $t = 1, \ldots, A!$. □

Hence, we consider the following hypotheses instead of (14),

$$\begin{aligned} &H_0 : 0 < \tau_1 = \cdots = \tau_A \\ &H_1 : 0 < \tau_1 < \cdots < \tau_A \\ &H_u : \tau_1, \ldots, \tau_A. \end{aligned} \tag{18}$$

Note that the inequality constraint $\tau > 0$ is also added to the equality constrained hypothesis H_0 for argument of symmetry. Using a uniform prior for $\boldsymbol{\tau}$ under the unconstrained hypothesis H_u, the balanced Bayes factor of the inequality constrained hypothesis H_1 against H_u equals

$$\begin{aligned} B_{1u} &= \frac{\int_{0<\tau_1<\cdots<\tau_A} \pi_0(\boldsymbol{\tau}|\mathbf{y}) \partial \boldsymbol{\tau}}{\int_{0<\tau_1<\cdots<\tau_A} \pi_0(\boldsymbol{\tau}) \partial \boldsymbol{\tau}} \\ &= A! \prod_{a=1}^{A} \frac{n_{\max,a}}{n_{\max,a} - 1} \\ &\quad \times \int_{0<\tau_1<\cdots<\tau_A} \pi_0(\tau_1, \ldots, \tau_A|\mathbf{y}) \partial \tau_1 \ldots \partial \tau_A \\ &\approx A! \prod_{a=1}^{A} \frac{n_{\max,a}}{n_{\max,a} - 1} S^{-1} \\ &\quad \times \sum_{s=1}^{S} I(0 < \tau_1^{(s)} < \cdots < \tau_A^{(s)}), \end{aligned} \tag{19}$$

where $\tau_a^{(s)}$ is the s-th posterior draw of τ_a.

3.1.1 Comparison with likelihood ratio tests

Although likelihood ratio tests are most often used, they may not be optimal when testing inequality constrained hypotheses such as $H_1 : 0 < \tau_1 < \cdots < \tau_A$ in (18). When hypotheses of different complexity levels are equally supported by the data, the likelihood ratio test will not prefer the simplest hypothesis. This is now illustrated.

Let the likelihood under H_1 be given by $g_1(\mathbf{y}|\boldsymbol{\tau}, \sigma^2, \mu) = g_u(\mathbf{y}|\boldsymbol{\tau}, \sigma^2, \mu) 1_{\boldsymbol{\tau} \in \mathbf{T}_1}(\boldsymbol{\tau})$, where $\mathbf{T}_1 = \{\boldsymbol{\tau} | 0 < \tau_1 < \cdots < \tau_A\}$ and $1_{\{\cdot\}}$ is the indicator function. Now, assume that the maximum likelihood estimates under both hypotheses are equal, i.e., $\hat{\boldsymbol{\tau}}_1 = \hat{\boldsymbol{\tau}}_u$. Consequently, the likelihood ratio statistic of H_1 against the unconstrained hypothesis H_u is equal to one when $\hat{\boldsymbol{\tau}}_1 = \hat{\boldsymbol{\tau}}_u$, which means that both hypotheses are equally supported by that data. This is a result of ignoring a difference in parsimony between the hypothesis H_1 and H_u. The more complex hypothesis H_1 should be preferred when there is enough additional supporting information. Otherwise, the simpler null hypothesis is to be preferred, which refers to Ockham's razor principle.

Furthermore, in order to obtain p-values for testing an inequality constrained hypothesis against the unconstrained hypothesis, so-called level probabilities must be computed which serve as chi-square weights. This is described by Silvapulle and Sen (2005) when testing inequalities on mean parameters. When testing inequality constraints on intraclass correlations such as H_1, such methods are not available.

The parsimony of H_1 and H_u is incorporated in the Bayes factor through the prior probabilities that the inequality constraints hold. When the likelihood is completely located in the inequality constrained space of H_1, this also holds for the posterior, and $\int_{0<\tau_1<\cdots<\tau_A} \pi_0(\boldsymbol{\tau}|\mathbf{y}) \partial \boldsymbol{\tau} = 1$. It follows that, the Bayes factor, defined in (19), supports H_1 against H_u with $B_{1u} = A! \prod_{a=1}^{A} \frac{n_{\max,a}}{n_{\max,a}-1} > 1$.

3.2 Bayes factor for equality constrained compound symmetry structures

A generalization can be made to test the equality of multiple (intra-class) parameters such as $H_0 : \tau_1 = \cdots = \tau_A$ against the unconstrained hypothesis H_u. This generalization leads to a Bayes factor with a comparable structure as in (17), which is proven in the following theorem.

Theorem 1 *Consider the selection problem of the equality constrained hypothesis* H_0, *where* $\tau = \tau_1 = \cdots = \tau_A$, *against the unconstrained hypothesis* H_u. *Let* $\pi_u(\boldsymbol{\tau}, \sigma^2, \mu)$ *denote a proper prior of* $(\boldsymbol{\tau}, \sigma^2, \mu)$ *under* H_u *and let* $\pi_0(\tau, \sigma^2, \mu) = \pi_u(\tau \mathbf{1}_A, \sigma^2, \mu)/c_0$ *under the equality constrained model with*

$$c_0 = \int_\tau \pi_u(\tau \mathbf{1}_A) d\tau. \tag{20}$$

Then, the Bayes factor can be expressed as

$$B_{0u} = \frac{\int_\tau \pi_u(\tau \mathbf{1}_A | \mathbf{y}) d\tau}{\int_\tau \pi_u(\tau \mathbf{1}_A) d\tau}. \tag{21}$$

Proof First note that $g_0(\mathbf{y} \mid \tau, \sigma^2, \mu) = g_u(\mathbf{y} \mid \tau \mathbf{1}_A, \sigma^2, \mu)$. Furthermore, when taking into account that $\pi_0(\tau, \sigma^2, \mu) = \pi_u(\tau \mathbf{1}_A, \sigma^2, \mu)/c_0$ where c_0 is defined in (20), the Bayes factor can be expressed as

$$\begin{aligned}
B_{0u} &= \frac{m_0(\mathbf{y})}{m_u(\mathbf{y})} \\
&= \frac{\int g_0(\mathbf{y} \mid \tau, \sigma^2, \mu) \pi_0(\tau, \sigma^2, \mu) d\mu d\sigma^2 d\tau}{\int g_u(\mathbf{y} \mid \boldsymbol{\tau}, \sigma^2, \mu) \pi_u(\boldsymbol{\tau}, \sigma^2, \mu) d\mu d\sigma^2 d\boldsymbol{\tau}} \\
&= \frac{\int g_u(\mathbf{y} \mid \tau \mathbf{1}_A, \sigma^2, \mu) \pi_u(\tau \mathbf{1}_A, \sigma^2, \mu)/c_0 d\mu d\sigma^2 d\tau}{\int g_u(\mathbf{y} \mid \boldsymbol{\tau}, \sigma^2, \mu) \pi_u(\boldsymbol{\tau}, \sigma^2, \mu) d\mu d\sigma^2 d\boldsymbol{\tau}} \\
&= \frac{\int \frac{g_u(\mathbf{y} | \tau \mathbf{1}_A, \sigma^2, \mu) \pi_u(\tau \mathbf{1}, \sigma^2, \mu)}{\int g_u(\mathbf{y} | \boldsymbol{\tau}, \sigma^2, \mu) \pi_u(\boldsymbol{\tau}, \sigma^2, \mu) d\mu d\sigma^2 d\boldsymbol{\tau}} d\mu d\sigma^2 d\tau}{c_0} \\
&= \frac{\int \pi_u(\tau \mathbf{1}_A, \sigma^2, \mu \mid \mathbf{y}) d\mu d\sigma^2 d\tau}{c_0} \\
&= \frac{f_0}{c_0},
\end{aligned} \tag{22}$$

where $f_0 = \int_\tau \pi_u(\tau \mathbf{1}_A | \mathbf{y}) d\tau$, which completes the proof. □

In expression (22), f_0 can be interpreted as a measure of fit and c_0 can be interpreted as a measure of complexity. Note that in the case of the inequality constrained hypothesis H_1, f_1 and c_1 were probabilities which is not the case in (22) where f_1 and c_1 can be seen as the surfaces under the posterior and prior density, respectively, on the line $\tau_1 = \cdots = \tau_A$. Therefore, the Bayes factor B_{0u} of $H_0 : \tau_1 = \cdots = \tau_A$ against $H_u : \tau_1, \ldots, \tau_A$ is unbounded which is not the case for $B_{1u} \leq c_1^{-1}$ of the inequality constrained hypothesis H_1 against H_u. For this reason, the unconstrained hypothesis can be seen as infinitely more complex in terms of allowed parameter space in comparison to an equality constrained hypothesis such as H_0.

When testing equality constrained hypotheses using the Bayes factor, it is important that a proper prior is chosen that is not too vague. The reason is that the Bayes factor for the equality constrained hypothesis against the unconstrained hypothesis becomes arbitrarily large when chosen the prior vaguely enough. This is also known as Bartlett's paradox (Bartlett 1957; Jeffreys 1961). In this setting, however, the parameters of interest $\boldsymbol{\tau}$ are bounded. For this reason, a uniform prior on $\boldsymbol{\tau}$ is a natural choice and the paradox is not an issue.

It is interesting to note the resemblance of (17) and (22) with the Savage-Dickey density ratio of the Bayes factor (Dickey 1971), i.e., $B_{0u} = \frac{\pi_0(\tau^* | \mathbf{y})}{\pi_0(\tau^*)}$, for testing $H_0 : \tau = \tau^*$ against $H_u : \tau$ which holds if $\pi_0(\mu, \sigma^2) = \pi_u(\mu, \sigma^2 \mid \tau = \tau^*)$. This was mentioned by Wetzels et al. (2010).

3.2.1 Numerical procedure for estimating B_{0u}

For the multivariate normal model with compound symmetry covariance structure (13), an explicit expression for the Bayes factor of $H_0 : \tau_1 = \cdots = \tau_A$ (or $H_0 : 0 < \tau_1 = \cdots = \tau_A$) versus $H_u : \tau_1, \ldots, \tau_A$ cannot be obtained because the marginal posterior of $\boldsymbol{\tau}$ does not have a common distribution.

For this reason, we propose a numerical procedure for estimating f_0 and c_0 in B_{0u} based on prior and posterior samples of the parameter vector $\boldsymbol{\tau}$, respectively, under the unconstrained hypothesis. Based on these samples, nonparametric multivariate density estimates (Scott 1992), such as multivariate histograms or multivariate kernel density estimation, can be used to approximate the marginal prior and posterior density of $\boldsymbol{\tau}$. Subsequently, the marginal posterior density can be estimated on the line $\tau \mathbf{1}_A$ by collecting pairs $(\boldsymbol{\tau}^{(s)}, \hat{\pi}_u(\boldsymbol{\tau}^{(s)} \mid \mathbf{y}))$ where $\hat{\pi}_u(\cdot | \mathbf{y})$ is the estimated posterior density, for some set of values $\boldsymbol{\tau}^{(s)} = \tau^{(s)} \mathbf{1}_A$, for $s = 1, \ldots, S$. Finally f_0 can be estimated as the surface of the cross section of the marginal posterior through the line $\tau \mathbf{1}_A$, i.e.,

$$\begin{aligned}
\hat{f}_0 &= \sum_{s=1}^{S-1} \hat{\pi}_u(0.5(\boldsymbol{\tau}^{(s+1)} + \boldsymbol{\tau}^{(s)}) \mid \mathbf{y}) \|\boldsymbol{\tau}^{(s+1)} - \boldsymbol{\tau}^{(s)}\|_E \\
&= \sqrt{A} \sum_{s=1}^{S-1} \hat{\pi}_u(0.5(\tau^{(s+1)} + \tau^{(s)}) \mathbf{1}_A \mid \mathbf{y}) \\
&\quad \times (\tau^{(s+1)} - \tau^{(s)}),
\end{aligned} \tag{23}$$

where $\| \cdot \|_E$ is the Euclidian norm. The estimate of c_1 can be obtained in a similar way using the estimated marginal prior density estimate denoted by $\hat{\pi}_u(\cdot)$.

Example 1 A data set was generated for $(\tau_1, \tau_2, \sigma^2, \mu) = (0.4, 0.5, 1, 0)$ with 6 groups of size $n_j = 6$ for sector $a = 1$

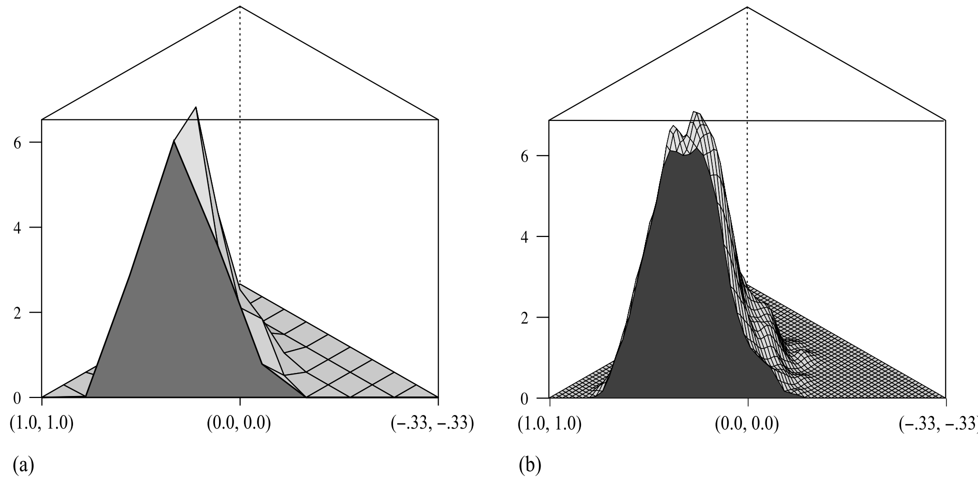

Fig. 4 Cross section on $\tau_1 = \tau_2$ of the unconstrained posterior sample density (obtained using kde2d of the R-program) for a posterior sample of (τ_1, τ_2) of size 10,000 by dividing the space of (τ_1, τ_2) in a 10×10 grid (**a**) and a 60×60 grid (**b**). The estimate $\hat{f}_0$ is computed as the surface of the cross section (2.78 and 2.82, respectively)

Table 1 Estimates of f_0 for different sample sizes and number of grid points

		Sample size		
		5,000	10,000	50,000
	10	2.83	2.78	2.71
Grid points	30	2.85	2.82	2.77
	60	2.85	2.82	2.76

and 6 groups of size $n_j = 4$ for sector $a = 2$. The Bayes factor was calculated of $H_0: \tau_1 = \tau_2$ against $H_u: \tau_1, \tau_2$ using the unconstrained prior (16) with $\pi_u(\tau_1, \tau_2) = \frac{3\cdot 5}{4\cdot 6} = 0.625$ for $(\tau_1, \tau_2) \in (-\frac{1}{5}, 1) \times (-\frac{1}{3}, 1)$. Hence, $c_0 = \sqrt{2}\frac{6}{5} \cdot 0.625 = 1.06$. The estimates of f_0 can be found in Table 1 for different numbers of grid points on the interval $(-\frac{1}{3}, 1)$ and different sample sizes. Figure 4 displays the estimate of f_0 as the surface of the cross section of the unconstrained posterior sample density estimate (obtained using the 2-dimensional kernel density estimator kde2d of the software package R) on the line $\tau_1 = \tau_2$ for the sample of size 10,000 and for 10 and 60 grid points, respectively. The Bayes factor is approximately $\hat{B}_{0u} = \frac{\hat{f}_0}{c_0} = 2.6$ which can be interpreted as small evidence in favor of H_0 against H_u.

4 Empirical data example: 1982 HSB survey data

The 1982 High School and Beyond (HSB, Raudenbush and Bryk 2002) Survey are considered, which include information on 7,185 students nested within 160 schools: 90 public and 70 Catholic. The sample from each school varies from a minimum of 14 to a maximum of 67 with an average of 45 students per school.

The outcome variable of interest was a standardized measure of math achievement. The observed score of student i in schools j is denoted as y_{ij}. A two-level multilevel model was defined to model the math scores: a student level and a school level. According to the study of Raudenbush and Bryk (2002), student's socioeconomic status (SES), school-average socioeconomic status (MSE), and the distinction in Catholic (SECTOR = 1) and public schools (SECTOR = 0) were important predictors in explaining variance in the math achievements. The following random intercept model was considered

$$\begin{aligned} y_{ij} &= \beta_{00} + \beta_{01}(\text{MSES})_j + \beta_{02}(\text{SECTOR})_j \\ &\quad + \beta_{10}((\text{SES})_{ij} - (\text{MSES})_j) \\ &\quad + \beta_{11}(\text{MSES})_j((\text{SES})_{ij} - (\text{MSES})_j) \\ &\quad + \beta_{12}(\text{SECTOR})_j((\text{SES})_{ij} - (\text{MSES})_j) \\ &\quad + u_{0j} + \varepsilon_{ij}, \\ &= \mathbf{x}'_{ij}\boldsymbol{\beta} + u_{0j} + \varepsilon_{ij} \end{aligned} \tag{24}$$

where $\boldsymbol{\beta} = (\beta_{00}, \beta_{01}, \beta_{02}, \beta_{10}, \beta_{11}, \beta_{12})'$, $\mathbf{x}_{ij}$ is a vector of length 6 containing the corresponding predictors. When using the parametrization of Σ_n^{CS1}, the random school effect is distributed as $u_{0j} \sim N(0, \tau\sigma^2)$, and the random error on the student level as $\varepsilon_{ij} \sim N(0, (1-\tau)\sigma^2)$.

Students are assumed to be nested within schools and the random intercept model in (24) assumes that the correlation between student-level math achievements is constant across public and catholic schools. The invariant correlation implies that the school's contribution to student-level math achievement is constant over public and catholic schools. This follows directly from the fact that the variance in math achievements is explained by an invariant between-school and between-student variance. Given socioeconomic differences, it is reasonable to assume that the average public school's contribution to student achievement differs from the average Catholic's contribution.

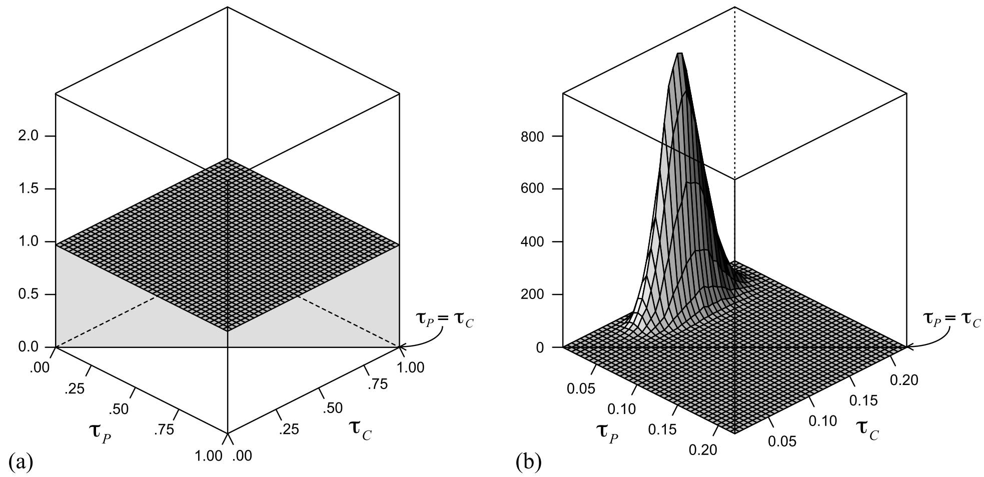

Fig. 5 Prior density (**a**) and approximated posterior density (**b**) of (τ_P, τ_C) for the HSB data. The surface of the cross section through the line $\tau_P = \tau_C$ (*grey area*) was equal to $\sqrt{2} \cdot 0.97 = 1.37$ and 0 for the prior and posterior, respectively

When assuming heterogenous intra-class correlations, $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_{160}')'$ are assumed to be multivariate normally distributed according to (13), such that

$$\mathbf{y} = N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathrm{diag}(\boldsymbol{\Sigma}_{\tau_1,n_1}, \ldots, \boldsymbol{\Sigma}_{\tau_J,n_J})), \tag{25}$$

where

$$\tau_j = \begin{cases} \tau_C & \text{if school } j \text{ is a Catholic school} \\ \tau_P & \text{if school } j \text{ is a public school,} \end{cases}$$

$\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_J')'$, $\mathbf{X}_j = (\mathbf{x}_{j1}, \ldots, \mathbf{x}_{j,n_j})'$, $\boldsymbol{\Sigma}_{\tau_j,n_j} = \tau_j \mathbf{J}_{n_j} + (1-\tau_j)\mathbf{I}_{n_j}$, with $\mathbf{J}_{n_j}$ a $n_j \times n_j$ matrix containing ones and $\mathbf{I}_{n_j}$ a $n_j \times n_j$ identity matrix.

Interest is focused on which of the following three model assumptions is most supported by the data,

$$\mathcal{M}_1 : \tau_C = \tau_P > 0$$

$$\mathcal{M}_2 : \tau_C > \tau_P > 0$$

$$\mathcal{M}_3 : \tau_P > \tau_C > 0.$$

In model $\mathcal{M}_1$, it is assumed that the math achievements within each school resemble each other to the same degree across public and catholic schools, conditional on explained differences by SES and MSES. In model $\mathcal{M}_2$, it is assumed that the student-level math achievements are stronger correlated within Catholic schools than within public schools. In model $\mathcal{M}_3$, the contrary of $\mathcal{M}_2$ is assumed.

A prior under the unconstrained model $\mathcal{M}_0$ is specified according to (16). Let $\mathcal{C}$ denote the set of indices $j = 1, \ldots, 160$. Then, $\max_{j \in \mathcal{C}}\{n_j\} = 67$ and $\max_{j \notin \mathcal{C}}\{n_j\} = 61$, which are both Catholic schools. Then, the following prior can be defined

$$\pi_0(\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) = \pi_0(\tau_C, \tau_P)\pi_0(\boldsymbol{\beta})\pi_0(\sigma^2),$$

where

$$\pi_0(\tau_C, \tau_P) = (1 + 1/66)^{-1}(1 + 1/60)^{-1} \times I(-1/66 < \tau_C < 1) I(-1/60 < \tau_P < 1)$$

$$\pi_0(\boldsymbol{\beta}) = \prod_{d=1}^{6} U(-1\mathrm{e}10, 1\mathrm{e}10)$$

$$\pi_0(\sigma^2) = \text{inv-}\chi^2(1, 1).$$

In Fig. 5a, the uniform prior of the intra-class correlations (τ_C, τ_P) is displayed in the region $(0,1) \times (0,1)$. Because the interest is in τ_C and τ_P which have proper noninformative priors, the hyperparameters are simply chosen such that they are relatively vague.

The likelihood of model (25) is given by

$$f(\mathbf{y} \mid \boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta}) \propto |\sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\left[\sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\tau}}\right]^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\tau}} = \mathrm{diag}(\boldsymbol{\Sigma}_{\tau_j,n_1}, \ldots, \boldsymbol{\Sigma}_{\tau_j,n_J})$. The joint posterior of the model parameters can be expressed as $\pi_0(\boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta}|\mathbf{y}) \propto f(\mathbf{y} \mid \boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta})\ \pi_0(\boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta})$.

An analytical expression of the marginal posterior of the intra-class correlations can be found by integrating out σ^2 and $\boldsymbol{\beta}$ of the joint posterior. In the Appendix it is shown that the resulting kernel of the marginal posterior of $\boldsymbol{\tau} = (\tau_C, \tau_P)'$ is given by

$$\pi(\boldsymbol{\tau}|\mathbf{y}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} |\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}|^{-\frac{1}{2}} \times \left(\mathbf{y}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{y} - \mathbf{y}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X})^{-1} \times \mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{y} + 1\right)^{-\frac{N-5}{2}}.$$

Although this posterior does not have a common form, samples can be obtained in a straightforward way using Metropolis-Hastings (Metropolis et al. 1953; Hastings 1970). The proposal density was a truncated normal with the mean equal to the previous draw and a standard deviation of 0.03 in the interval $(-1/66, 1)$ for τ_C and $(-1/60, 1)$ for τ_P. A graphical representation of the posterior sample can be found in Fig. 5b.

Based on the posterior sample, the posterior probability of $H_1 : \tau_C, \tau_P > 0$ was estimated as

$$\begin{aligned}P(H_1 \mid \mathbf{y}) &= P(\tau_C, \tau_P > 0|\mathbf{y}) \\ &= \int_{\tau_C,\tau_P>0} \pi(\tau_C, \tau_P|\mathbf{y})\partial\tau_C\partial\tau_P \\ &\approx S^{-1}\sum_s I\big(\tau_C^{(s)}, \tau_P^{(s)} > 0\big) = 1.\end{aligned}$$

This suggests that student achievements are positively correlated within schools, with invariant intra-class correlations over public and Catholic schools, given differences in socio-economic status. Furthermore, the posterior estimates of the fixed effects β did not change when fitting this model with separate intra-class correlations τ_C and τ_p instead of using a single $\tau = \tau_C = \tau_p$.

The justification of an invariant intra-class correlation is tested using Bayes factors, which are estimated using the uniform prior density and posterior sample of (τ_P, τ_C) according to the methodology described in Sect. 3. This yields the following Bayes factors,

$$\begin{aligned}\hat{B}_{10} &= \frac{0}{\sqrt{2}(1+1/60)^{-1}(1+1/66)^{-1}} = 0 \\ \hat{B}_{20} &= \frac{S^{-1}\sum_s I(\tau_C^{(s)} > \tau_P^{(s)} > 0)}{0.5(1+1/60)^{-1}(1+1/66)^{-1}} = \frac{1}{0.4844629} \\ &= 2.06 \\ \hat{B}_{30} &= \frac{S^{-1}\sum_s I(\tau_P^{(s)} > \tau_C^{(s)} > 0)}{0.5(1+1/60)^{-1}(1+1/66)^{-1}} = \frac{0}{0.4844629} = 0.\end{aligned}$$

Consequently, the Bayes factors between the constrained models were given by $\hat{B}_{21} = \hat{B}_{20}/\hat{B}_{10} = \infty$ and $\hat{B}_{23} = \hat{B}_{20}/\hat{B}_{30} = \infty$, from which it can be concluded that the evidence for model $\mathcal{M}_2$ against both $\mathcal{M}_1$ and $\mathcal{M}_3$ is overwhelming. Hence, there is decisive evidence that the intra-class correlation is significantly larger in Catholic schools in comparison to public schools, given socio-economic differences. Furthermore, the 90% credible intervals of τ_C and τ_p were equal to $(0.11, 0.17)$ and $(0.03, 0.05)$, respectively. It can therefore be concluded that there is hardly any within-school clustering of students in public schools but a strong clustering in Catholic schools. As a result, the school's contribution to the student-level math performance is significantly higher for Catholic schools than for public schools. For the public schools, the total variance in the student-level math achievements is almost equal to within-school residual variance since the between-school variance is very small.

5 Summary

Bayesian tests on covariance parameter(s) of a compound symmetry structure in the marginal model are proposed, where observations are multivariate normally distributed. The outcome of the test can be used in model building to justify any conditional independence assumptions from a marginal modeling framework. In our approach, innovative proper prior distributions were introduced for the variance components, such that the positive definiteness of the (compound symmetry) covariance matrix was ensured.

A generalization was made to Bayes factors for testing heterogenous compound symmetry structures, where equality or inequality constrained hypotheses of covariance parameters were considered. In the case of inequality constrained hypotheses, it was shown that the proposed priors defined a balanced Bayes factor, which correctly balances between model fit and model complexity. The prior and posterior samples can be used to obtain the Bayes factor for testing equality of the parameter vector of interest. This latter procedure was used for testing heterogeneity of the intra-class correlation of public and Catholic schools for the 1982 High School and Beyond Survey data. It was concluded that the intra-class correlation between Catholic schools was larger than for public schools, and therefore, the use of a conditional model with a homogeneous intra-class correlation for all school types is not advisable.

Bayesian tests were developed for normally distributed observations. However, the compound symmetry covariance structure can also be defined for non-normally distributed categorical data within the class of generalized linear models. When a normal underlying augmented variable can be defined, the proposed Bayesian tests can be defined conditional on the augmented variable. To integrate over the augmented data the MCMC algorithm needs to be extended with an additional step. As a result, the compound symmetry covariance structure can be tested with the proposed Bayesian tests for categorical data.

Although this paper focused on a compound symmetry covariance structure, similar tests can be constructed for variance components in other structured covariance matrices, such as auto-regressive or Toeplitz. A challenge here is the specification of proper prior distributions for the variance components that result in a positive definite covariance matrices with the appropriate structure. When this can be achieved, effective and flexible statistical tools can be designed for testing complex dependency structures.

Appendix: Derivation of the marginal posterior of $\boldsymbol{\Sigma}_{\boldsymbol{\tau}}$

The model is given by

$$\mathbf{y}_j = \mathcal{N}(\mathbf{X}_j\boldsymbol{\beta}, \sigma^2\boldsymbol{\Sigma}_{\tau_j,n_j}), \tag{26}$$

with

$$\boldsymbol{\Sigma}_{\tau_j,n_j} = \begin{bmatrix} 1 & \tau_j & \cdots & \tau_j \\ \tau_j & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \tau_j \\ \tau_j & \cdots & \tau_j & 1 \end{bmatrix} = (1-\tau_j)\mathbf{I}_{n_j} + \tau_j\mathbf{J}_{n_j},$$

where τ_j is the intraclass correlation in group j, and $\boldsymbol{\beta}$ contain the D fixed effects. Furthermore,

$$\mathbf{y} = \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{\tau}}), \tag{27}$$

with $\boldsymbol{\Sigma}_{\boldsymbol{\tau}} = \text{diag}(\boldsymbol{\Sigma}_{\tau_1,n_1}, \ldots, \boldsymbol{\Sigma}_{\tau_J,n_J})$, $\mathbf{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_J')'$ and $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_J')'$. The likelihood is given by

$$f(\mathbf{y}|\boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta}) \propto |\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\left[\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{\tau}}\right]^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}.$$

The following noninformative proper prior is used under the unconstrained model $\mathcal{M}_0 : \tau_1, \ldots, \tau_A$,

$$\pi_0(\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma^2) = \pi_0(\boldsymbol{\tau})\pi_0(\boldsymbol{\beta})\pi_0(\sigma^2), \quad \text{where}$$
$$\pi_0(\boldsymbol{\tau}) = \prod_{a=1}^{A} \pi_0(\tau_a) = \prod_{a=1}^{A} U\left(-\left(\max_{j\in\mathcal{J}_a}\{n_j\} - 1\right)^{-1}, 1\right)$$
$$\pi_0(\boldsymbol{\beta}) = \prod_{d=1}^{D} U(-1\text{e}\ell, 1\text{e}\ell)$$
$$\pi_0(\sigma^2) = \text{inv-}\chi^2(1,1), \tag{28}$$

where $\mathcal{J}_a$ is the set of group numbers belonging to sector $a = 1, \ldots, A$. The bound ℓ in the multivariate uniform prior of the nuisance parameter vector $\boldsymbol{\beta}$ is chosen such the likelihood is essentially zero outside $(-1\text{e}\ell, 1\text{e}\ell)^D$, e.g. $\ell = 1\text{e}10$. Consequently, the joint posterior is essentially proportional to $\pi_0(\sigma^2) f(\mathbf{y}|\boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta})$, i.e.,

$$\pi(\boldsymbol{\tau}, \sigma^2, \boldsymbol{\beta}|\mathbf{y}) \propto (\sigma^2)^{-\frac{N+1}{2}-1}|\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2\sigma^2}((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+1)\right\}.$$

First, σ^2 is integrated out of $\pi(\tau, \sigma^2, \boldsymbol{\beta}|\mathbf{y})$ using a scaled inv-χ^2 distribution for σ^2 with hyper-parameters

$$\nu_{\sigma^2} = N+1$$
$$s^2_{\sigma^2} = (N+1)^{-1}((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+1),$$

by

$$\begin{aligned}\pi(\boldsymbol{\tau}, \boldsymbol{\beta}|\mathbf{y}) &\propto s_{\sigma^2}^{-\nu_{\sigma^2}} \frac{\Gamma(\nu_{\sigma^2}/2)}{(\nu_{\sigma^2}/2)^{\nu_{\sigma^2}/2}}|\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} \int_{\sigma^2} \frac{(\nu_{\sigma^2}/2)^{\nu_{\sigma^2}/2}}{\Gamma(\nu_{\sigma^2}/2)} \\ &\quad \times s_{\sigma^2}^{\nu_{\sigma^2}}(\sigma^2)^{-\nu_{\sigma^2}/2-1}\exp\left\{-\frac{\nu_{\sigma^2}s^2_{\sigma^2}}{2\sigma^2}\right\}d\sigma^2 \\ &\propto |\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}} s_{\sigma^2}^{-\nu_{\sigma^2}} \\ &\propto |\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}}\left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+1\right)^{-\frac{N+1}{2}}.\end{aligned}$$

When denoting

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{y}$$
$$\mathbf{V}_{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}\right)^{-1}$$
$$\xi = \mathbf{y}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{V}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\beta}} + 1,$$

we can rewrite

$$\begin{aligned}&\left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+1\right)^{-\frac{N+1}{2}} \\ &\quad = \left(\xi + (\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right)^{-\frac{N+1}{2}} \\ &\quad = \xi^{-\frac{N+1}{2}}(1+\xi^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\mathbf{V}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}))^{-\frac{N+1}{2}}.\end{aligned}$$

Furthermore, let

$$\mathbf{S}_{\boldsymbol{\beta}} = (N-K+1)^{-1}\xi\mathbf{V}_{\boldsymbol{\beta}}$$
$$\mathbf{m}_{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$$
$$\nu_{\boldsymbol{\beta}} = N-K+1,$$

which are the hyperparameters of the multivariate t distribution of $\boldsymbol{\beta}$ which we can integrate out of $\pi(\tau, \boldsymbol{\beta}|\mathbf{y})$. This results in

$$\begin{aligned}\pi(\boldsymbol{\tau}) &\propto |\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}}\xi^{-\frac{N+1}{2}}\frac{\Gamma(\frac{\nu_{\boldsymbol{\beta}}}{2})\nu_{\boldsymbol{\beta}}^{K/2}\pi^{K/2}}{\Gamma(\frac{\nu_{\boldsymbol{\beta}}+K}{2})}|\mathbf{S}_{\boldsymbol{\beta}}|^{\frac{1}{2}} \\ &\quad \times \int_{\boldsymbol{\beta}} \frac{\Gamma(\frac{\nu_{\boldsymbol{\beta}}+K}{2})}{\Gamma(\frac{\nu_{\boldsymbol{\beta}}}{2})\nu_{\boldsymbol{\beta}}^{K/2}\pi^{K/2}}|\mathbf{S}_{\boldsymbol{\beta}}|^{-\frac{1}{2}} \\ &\quad \times \left(1+\frac{1}{\nu_{\boldsymbol{\beta}}}(\boldsymbol{\beta}-\mathbf{m}_{\boldsymbol{\beta}})'\mathbf{S}_{\boldsymbol{\beta}}^{-1}(\boldsymbol{\beta}-\mathbf{m}_{\boldsymbol{\beta}})\right)^{-\frac{\nu_{\boldsymbol{\beta}}+K}{2}}\partial\boldsymbol{\beta} \\ &\propto |\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}}\xi^{-\frac{N+1}{2}}|(N-K+1)^{-1}\xi\mathbf{V}_{\boldsymbol{\beta}}|^{\frac{1}{2}}\end{aligned}$$

$$\propto |\boldsymbol{\Sigma}_{\boldsymbol{\tau}}|^{-\frac{1}{2}}\big(\mathbf{y}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{y}-\mathbf{y}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}\big(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}\big)^{-1} \times \mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{y}+1\big)^{-\frac{N-K+1}{2}}|\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1}\mathbf{X}|^{-\frac{1}{2}},$$

where

$$|\boldsymbol{\Sigma}_{\boldsymbol{\tau}}| = \prod_{j=1}^{J}|\boldsymbol{\Sigma}_{\tau_j}| = \prod_{j=1}^{J}((n_j-1)\tau_j+1)(1-\tau_j)^{n_j-1}$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\tau}}^{-1} = \mathrm{diag}(\boldsymbol{\Sigma}_{\tau_1,n_1}^{-1},\ldots,\boldsymbol{\Sigma}_{\tau_J,n_J}^{-1})$$

with

$$\boldsymbol{\Sigma}_{\tau_j,n_j}^{-1} = \zeta_{I,n_j}\mathbf{I}_{n_j}+\zeta_{J,n_j}\mathbf{J}_{n_j} \quad \text{with}$$

$$\zeta_{I,n_j} = \frac{1+(n_j-1)\tau_j}{1+(n_j-2)\tau_j-(n_j-1)\tau_j^2} = (1-\tau_j)^{-1}$$

$$\zeta_{J,n_j} = \frac{-\tau_j}{1+(n_j-2)\tau_j-(n_j-1)\tau_j^2} = \frac{-\tau_j}{(1-\tau_j)(1+(n_j-1)\tau_j)}.$$

References

Andrieu, C., Thomas, J.: A tutorial on adaptice MCMC. Stat. Comput. **18**, 343–373 (2008)

Bartlett, M.: A comment on Lindley's statistical paradox. Biometrika **44**, 533–534 (1957)

Berger, J.O., Bernardo, J.M.: On the development of the reference prior method. In: Bernardo, J.M., Berger, J.O., Smith, A.F.M. (eds.) Bayesian Statistics, vol. 4, pp. 35–600. London: Oxford Univ. Press (1992)

Berger, J.O., Mortera, J.: Default Bayes Factors for nonnested hypothesis testing. J. Am. Stat. Assoc. **94**, 542–554 (1999)

Berger, J.O., Pericchi, L.R.: The intrinsic Bayes factor for model selection and prediction. J. Am. Stat. Assoc. **91**, 109–122 (1996)

Box, G.E.P., Tiao, G.C.: Bayesian Inference in Statistical Analysis. Addison-Wesley, Reading (1973)

Browne, W.J.: MCMC algorithms for constrained variance matrices. Comput. Stat. Data Anal. **50**, 1655–1677 (2006)

Chib, S., Greenberg, E.: Understanding the Metropolis-Hastings algorithm. Am. Stat. **49**, 327–335 (1995)

Chib, S., Greenberg, E.: Analysis of multivariate probit models. Biometrika **85**, 347–361 (1998)

Chib, S., Jeliazkov, I.: Marginal likelihoods from the Metropolis-Hastings Output. J. Am. Stat. Assoc. **96**, 270–281 (2001)

Chung, Y., Dey, D.K.: Bayesian approach to estimation of intraclass correlation using reference prior. Commun. Stat. **26**, 2241–2255 (1998)

Dickey, J.: The weighted likelihood ratio linear hypotheses on normal location parameters. Ann. Stat. **42**, 204–223 (1971)

Fox, J.-P.: Bayesian Item Response Modeling: Theory and Applications. Springer, New York (2010)

Gelfand, A.E., Hills, S.E., Racine-Poon, A., Smith, A.F.M.: Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Am. Stat. Assoc. **85**, 972–985 (1990)

Gelman, A., Hill, J.: Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York (2007)

Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97–109 (1970)

Jeffreys, H.: Theory of Probability, 3rd edn. Oxford University Press, New York (1961)

Kass, R.E., Raftery, A.E.: Bayes factors. J. Am. Stat. Assoc. **90**, 773–795 (1995)

Kato, B., Hoijtink, H.: Testing homogeneity in a random intercept model using asymptotic posterior predictive and plug-in p-values. Stat. Neerl. **58**, 179–196 (2004)

Klugkist, Hoijtink: The Bayes factor for inequality and about equality constrained models. Comput. Stat. Data Anal. **51**, 6367–6379 (2007)

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equations of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)

Mulder, J., Hoijtink, H., Klugkist, I.: Equality and inequality constrained multivariate linear models: objective model selection using constrained posterior priors. J. Stat. Plan. Inference **140**, 887–906 (2010)

O'Hagen, A.: Fractional Bayes factors for model comparison (with discussion). J. R. Stat. Soc. Ser. B **57**, 99–138 (1995)

Raudenbush, S.W., Bryk, A.S.: Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd edn. Sage, Thousand Oaks (2002)

Roberts, G.O., Sahu, S.K.: Updating schemes correlation structure, blocking and parameterization for the Gibbs sampler. J. R. Stat. Soc. B **59**, 291–317 (1997)

Scott, D.W.: Multivariate Density Estimation: Theory, Practice, and Visualization. Wiley, New York (1992)

Silvapulle, M.J., Sen, P.K.: Constrained Statistical Inference: Inequality, order, and Shape Restrictions. Wiley, Hoboken (2005)

Snijders, T.A.B., Bosker, R.J.: Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Sage, London (1999)

Szatkowski, A.: On the dynamic spaces and on the equations of motion of non-linear. Int. J. Circuit Theory Appl. **10**, 99–122 (1982)

Verbeke, G., Spiessens, B., Lesaffre, E.: Conditional linear mixed models. Am. Stat. **55**, 24–34 (2001)

Votaw, D.F.: Testing compound symmetry in a normal multivariate distribution. Ann. Math. Stat. **19**, 447–473 (1948)

Wetzels, R., Grasman, R.P.P.P., Wagenmakers, E.-J.: An encompassing prior generalization of the Savage-Dickey density ratio. Comput. Stat. Data Anal. **54**, 2094–2102 (2010)