



BIEMS: A Fortran 90 Program for Calculating Bayes Factors for Inequality and Equality Constrained Models

Joris Mulder
Tilburg University

Herbert Hoijtink
Utrecht University

Christiaan de Leeuw
Utrecht University

Abstract

This paper discusses a Fortran 90 program referred to as **BIEMS** (Bayesian inequality and equality constrained model selection) that can be used for calculating Bayes factors of multivariate normal linear models with equality and/or inequality constraints between the model parameters versus a model containing no constraints, which is referred to as the unconstrained model. The prior that is used under the unconstrained model is the conjugate expected-constrained posterior prior and the prior under the constrained model is proportional to the unconstrained prior truncated in the constrained space. This results in Bayes factors that appropriately balance between model fit and complexity for a broad class of constrained models. When the set of equality and/or inequality constraints in the model represents a hypothesis that applied researchers have in, for instance, (M)AN(C)OVA, (multivariate) regression, or repeated measurements, the obtained Bayes factor can be used to determine how much evidence is provided by the data in favor of the hypothesis in comparison to the unconstrained model. If several hypotheses are under investigation, the Bayes factors between the constrained models can be calculated using the obtained Bayes factors from **BIEMS**. Furthermore, posterior model probabilities of constrained models are provided which allows the user to compare the models directly with each other.

Keywords: Bayes factors, equality and inequality constrained models, Fortran 90, Gibbs sampler.

1. Introduction

This paper discusses a program that can be used to determine which model (or hypothesis) receives most evidence from the data given a set of models that contain equality constraints and/or inequality constraints on the model parameters. This will be done using the Bayes

factor, a Bayesian model selection criterion. The program **BIEMS** (Bayesian inequality and equality constrained model selection) is written in Fortran 90 using IMSL subroutines (Visual Numerics 2003). A user friendly interface has been developed for practitioners who are not familiar with the technical details of statistical modeling. The interface allows one to simply specify equality and inequality constraints on the parameters of interest. From the website <http://tinyurl.com/informativehypotheses>, it is possible to download a compiled version of the program, including the interface. Source codes are available along with this manuscript.

The program is designed for applied researchers, e.g., social or medical, who have specific expectations that can be formulated in terms of equality and inequality constraints between the model parameters. Hoijsink *et al.* (2008) refer to such expectations as informative hypotheses in contrast to hypotheses that do not contain any information about the relationship between the model parameters, such as the standard null and alternative in classical hypothesis testing. For references about the use of Bayes factors to test scientific theories, see Kass and Raftery (1995) and the references therein. The methodology that is used in **BIEMS** is based on Mulder *et al.* (2010) and Mulder *et al.* (2009). The first reference contains the technical details of the method, such as, prior specification. The latter focusses on repeated measurements and is more accessible for applied researchers.

In Section 2.1, the multivariate normal linear model is discussed. Section 2.2 describes how time-varying covariates can be included in the model. The formulation of model constraints is addressed in Section 2.3. Section 3 describes the specification of the unconstrained default prior which is the conjugate expected-constrained posterior prior (CECPP). In Section 3.1, two basic properties of this prior are discussed. Section 3.2 describes how linear restrictions are specified to obtain so-called constrained posterior priors. Subsequently, the hyper-parameters of the constrained posterior priors are calculated in Section 3.3 and the construction of the CECPP is given in Section 3.4 including the corresponding posterior distribution. In Section 3.5, it is discussed how minimal training samples are sampled which are used to construct the constrained posterior priors. In Section 4.1, it is described how the Bayes factor is estimated using MCMC (Markov chain Monte Carlo) sampling for an inequality constrained model versus the unconstrained model. Section 4.2 illustrates how the estimation error of the Bayes factor is determined. In Section 4.3, the procedure is discussed for estimating Bayes factors of models containing equality constraints versus the unconstrained model. Section 5 provides several short examples of model implementations for analysis of variance (ANOVA), linear regression, multivariate analysis of variance (MANOVA), and repeated measures. In Section 6, a short conclusion is given.

2. Constrained model specification

2.1. The multivariate normal linear model

The multivariate normal linear model can be written as

$$\mathbf{y}_i = \mathbf{M}^\top \mathbf{d}_i + \mathbf{A}^\top \mathbf{x}_i + \boldsymbol{\varepsilon}_i. \quad (1)$$

In this model, a P -variate dependent variable \mathbf{y}_i , $i = 1, \dots, N$, is explained by a vector \mathbf{d}_i of

size J containing group membership, i.e.,

$$d_{ij} = \begin{cases} 1 & \text{if the } i\text{th observation belongs to group } j \\ 0 & \text{otherwise,} \end{cases}$$

a K -variate explanatory variable \mathbf{x}_i , and ε_i a multivariate Gaussian error with $\varepsilon_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Furthermore, \mathbf{M} is a $J \times P$ matrix with group means where the (j, p) th element μ_{jp} denotes the mean (or intercept) of the p th dependent variable for the j th group and \mathbf{A} is a $K \times P$ matrix with regression coefficients where the (k, p) th element α_{kp} denotes the coefficient of the k th covariate for the p th dependent variable.

For N observations, the model can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}, \quad (2)$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_N^\top \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{d}_1^\top & \mathbf{x}_1^\top \\ \vdots & \vdots \\ \mathbf{d}_N^\top & \mathbf{x}_N^\top \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \mathbf{M} \\ \mathbf{A} \end{bmatrix} \quad (3)$$

with

$$\mathbf{M} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1P} \\ \vdots & \ddots & \vdots \\ \mu_{J1} & \cdots & \mu_{JP} \end{bmatrix}, \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} \alpha_{11} & \cdots & \alpha_{1P} \\ \vdots & \ddots & \vdots \\ \alpha_{K1} & \cdots & \alpha_{KP} \end{bmatrix}. \quad (4)$$

The likelihood of the multivariate normal linear model (2) is given by

$$\begin{aligned} g(\mathbf{Y}|\mathbf{X}, \mathbf{B}, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-N/2} \exp\left\{-\frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\mathbf{B})^\top(\mathbf{Y} - \mathbf{X}\mathbf{B})\right\} \\ &= \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top [\boldsymbol{\Sigma} \otimes (\mathbf{X}^\top \mathbf{X})^{-1}]^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\ &\quad \cdot |\boldsymbol{\Sigma}|^{-N/2} \exp\left\{-\frac{1}{2} \text{tr} \boldsymbol{\Sigma}^{-1} \mathbf{S}\right\} \\ &\propto \mathcal{N}_{\boldsymbol{\beta}|\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}, \boldsymbol{\Sigma} \otimes (\mathbf{X}^\top \mathbf{X})^{-1}) \text{inv-}\mathcal{W}_{\boldsymbol{\Sigma}}(\mathbf{S}, N), \end{aligned} \quad (5)$$

where $\boldsymbol{\beta}$ is a vector of length $D = P(J + K)$, which is the vectorization of \mathbf{B} , i.e., $\text{vec}(\mathbf{B}) = \boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}$, with

$$\hat{\mathbf{B}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

and

$$\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}). \quad (6)$$

The least squares estimate of $\boldsymbol{\Sigma}$, which is provided in the output of **BIEMS**, equals $\hat{\boldsymbol{\Sigma}} = \mathbf{S}/(N - J - K)$. Expression (5) shows that the likelihood is proportional to a normal-inverse-Wishart distribution. For this reason, a multivariate normal-inverse Wishart prior for $\{\boldsymbol{\beta}, \boldsymbol{\Sigma}\}$ is used in **BIEMS** which is generalized natural conjugate (Press 2005, p. 253–256), i.e., $\pi_0(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \pi_0(\boldsymbol{\beta})\pi_0(\boldsymbol{\Sigma})$. The unconstrained prior is discussed in Section 3.

2.2. Modeling time-varying covariates

When modeling repeated measurements using the multivariate normal linear model, time-varying covariates can be included in (2) according to

$$\mathbf{X} = \begin{bmatrix} \mathbf{d}_1^\top & \mathbf{x}_1^\top & \mathbf{w}_{11}^\top & \cdots & \mathbf{w}_{1U}^\top \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{d}_N^\top & \mathbf{x}_N^\top & \mathbf{w}_{N1}^\top & \cdots & \mathbf{w}_{NU}^\top \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} \mathbf{M} \\ \mathbf{A} \\ \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_U \end{bmatrix}, \quad (7)$$

with \mathbf{w}_{iu} is a vector of length P containing the scores on the u th time-varying covariates of the i th observation, $\mathbf{G}_u = \text{diag}(\gamma_{u1}, \dots, \gamma_{uP})$, where γ_{up} is the regression parameter of the u th time-varying covariate measured at the p th time point, for $u, = 1, \dots, U$. The matrices \mathbf{G}_u are diagonal to ensure that the u th time-varying covariate measured at time point p is not regressed on the outcome variable measured at time point $p^\top \neq p$. In the vectorization of \mathbf{B} , i.e., $\boldsymbol{\beta}$, which is important when specifying the (in)equality constraints of the model in (10), the zeros are omitted (see also Section 5.5). Note here that it is also possible to model more than one dependent variable as repeated measurement.

When time-varying covariates are present, the likelihood of the parameters in the structured parameter matrix \mathbf{B} given $\boldsymbol{\Sigma}$ can be obtained as follows. First, an unstructured parameter matrix is substituted for \mathbf{B} , which is denoted by

$$\mathbf{B}^* = \begin{bmatrix} \mathbf{M} \\ \mathbf{A} \\ \mathbf{G}_1^* \\ \vdots \\ \mathbf{G}_U^* \end{bmatrix},$$

where \mathbf{G}_u^* , for $u = 1, \dots, U$, are unstructured parameter matrices which replace the diagonal matrices \mathbf{G}_u , for $u = 1, \dots, U$, \mathbf{B} in (7). The vector with all off-diagonal parameters of the matrices \mathbf{G}_u^* will be denoted by $\boldsymbol{\gamma}^*$ which contains $\frac{1}{2}UP(P-1)$ elements. Therefore, the vectorization of \mathbf{B}^* is a permutation of the parameter vector $(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^{*\top})^\top$, where $\boldsymbol{\beta}$ is the vectorization of \mathbf{B} in (7) with zeros omitted. The likelihood of the unstructured parameter matrix \mathbf{B}^* and $\boldsymbol{\Sigma}$ is given in (5), and therefore, the likelihood of $\boldsymbol{\beta}, \boldsymbol{\gamma}^* | \boldsymbol{\Sigma}$ is multivariate normal. Subsequently, the likelihood of $\boldsymbol{\beta} | \boldsymbol{\Sigma}, \boldsymbol{\gamma}^* = \mathbf{0}$ is also multivariate normal and given by

$$\mathcal{N}(\hat{\boldsymbol{\beta}} - \text{cov}(\boldsymbol{\gamma}^*, \boldsymbol{\beta})\text{var}(\boldsymbol{\gamma}^*)^{-1}\hat{\boldsymbol{\gamma}}^*, \text{var}(\boldsymbol{\beta}) + \text{cov}(\boldsymbol{\gamma}^*, \boldsymbol{\beta})\text{var}(\boldsymbol{\gamma}^*)^{-1}\text{cov}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*)),$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}^*$ are the maximum likelihood estimates, and $\text{cov}(\boldsymbol{\gamma}^*, \boldsymbol{\beta})$, $\text{cov}(\boldsymbol{\beta}, \boldsymbol{\gamma}^*)$, $\text{var}(\boldsymbol{\beta})$, and $\text{var}(\boldsymbol{\gamma}^*)$ are the appropriate submatrices of the covariance matrix $\boldsymbol{\Sigma} \otimes (\mathbf{X}^\top \mathbf{X})^{-1}$. Subsequently, the conditional posterior of $\boldsymbol{\beta} | \boldsymbol{\Sigma}, \boldsymbol{\gamma}^* = \mathbf{0}$ can be obtained by combining this likelihood with the multivariate normal prior of $\boldsymbol{\beta}$ which is discussed in the next section.

An example of a model including a time-varying covariate is given in Section 5.5. More details about modeling time-varying covariates using the multivariate normal linear model can be found in Mulder *et al.* (2009).

2.3. Equality and inequality constrained multivariate normal linear models

The system of L_t inequality constraints on the parameters $\boldsymbol{\beta}$ of the inequality constrained model \mathcal{M}_t can be written as

$$\mathbf{R}_t \boldsymbol{\beta} > \mathbf{r}_t, \quad (8)$$

where \mathbf{R}_t is a $L_t \times D$ matrix containing inequality coefficients and \mathbf{r}_t is a vector of length L_t of constants. The augmented matrix $[\mathbf{R}_t | \mathbf{r}_t]$ should be implemented in the file `inequality_constraints.txt` in the stand-alone version of **BIEMS** (Appendix B). Examples are provided in Section 5.

For an equality constrained model \mathcal{M}_t , the system of equalities is given by

$$\mathbf{Q}_t \boldsymbol{\beta} = \mathbf{q}_t, \quad (9)$$

where \mathbf{Q}_t is a $H_t \times D$ matrix containing H_t equality coefficients and \mathbf{q}_t is a vector of length H_t containing constants. The augmented matrix $[\mathbf{Q}_t | \mathbf{q}_t]$ should be given in the file `equality_constraints.txt` (Appendix B).

When combining (8) and (9), we obtain a constrained model with inequalities and equalities. The system of constraints of model \mathcal{M}_t will be denoted by

$$\begin{bmatrix} \mathbf{R}_t \\ \mathbf{Q}_t \end{bmatrix} \boldsymbol{\beta} \geq_t \begin{bmatrix} \mathbf{r}_t \\ \mathbf{q}_t \end{bmatrix} \Leftrightarrow \tilde{\mathbf{R}}_t \boldsymbol{\beta} \geq_t \tilde{\mathbf{r}}_t, \quad (10)$$

where \geq_t denotes that each constraint of this system is either a “>” or a “=”, depending on the model \mathcal{M}_t .

3. Default prior specification

In order to compute Bayes factors, priors distributions must be specified for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. Because all constrained models are nested in the unconstrained model $\mathcal{M}_0 : \boldsymbol{\beta} \in R^D$, one only needs to specify a prior under the unconstrained model and use truncations of this prior under the constrained models, i.e., $\pi_t(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = c_t^{-1} \pi_0(\boldsymbol{\beta}, \boldsymbol{\Sigma}) 1_{\mathcal{M}_t}(\boldsymbol{\beta})$ with $c_t = \iint_{\boldsymbol{\beta} \in \mathcal{M}_t} \pi_0(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \partial \boldsymbol{\beta} \partial \boldsymbol{\Sigma}$, where π_0 denotes the prior under \mathcal{M}_0 and π_t denotes the prior under \mathcal{M}_t . This prior choice has two important advantages. First, only a single prior π_0 needs to be specified. Second, the Bayes factor B_{t0} of a constrained model against the unconstrained model can be obtained without computing the marginal likelihoods of the models. This is discussed in the Section 4. In this section, it will be described how a default prior under the unconstrained model can be obtained which results in Bayes factors that are effective for evaluating models containing equality and inequality constraints on the parameters.

3.1. Properties of the unconstrained prior

In Mulder *et al.* (2010) and Mulder *et al.* (2009), a prior specification was proposed that resulted in Bayes factors with an appropriate Occam’s razor for testing equality and inequality constrained models due to an effective balance between model fit and model complexity. The implementation of this method in **BIEMS** is discussed in this section.

The CECPP is used as unconstrained prior for $\boldsymbol{\beta}$ (Mulder *et al.* 2009). The CECPP is based on constrained posterior priors, which are obtained as follows. For the covariance matrix $\boldsymbol{\Sigma}$,

Jeffreys' prior is used, i.e., $\pi_0(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(P+1)/2}$. This noninformative improper prior can be used because of arguments of orthogonal transformations of scale and location parameters (Jeffreys 1961).

The CECPP of β has the following three key properties.

1. It is based on minimal training samples which are subsets of the data that contain minimal information to obtain a proper posterior prior (e.g., Berger and Pericchi 1996). This property is essential for testing models with equality constraints on the model parameters. The Bartlett paradox (Bartlett 1957) is then avoided.
2. It is centered on the boundary of the constrained parameter space under investigation. The boundary of the constrained space under investigation is defined as

$$\mathcal{B} = \{\beta | \tilde{\mathbf{R}}_0 \beta = \tilde{\mathbf{r}}_0\} \text{ with } \tilde{\mathbf{R}}_0 = \begin{bmatrix} \tilde{\mathbf{R}}_1 \\ \vdots \\ \tilde{\mathbf{R}}_T \end{bmatrix} \text{ and } \tilde{\mathbf{r}}_0 = \begin{bmatrix} \tilde{\mathbf{r}}_1 \\ \vdots \\ \tilde{\mathbf{r}}_T \end{bmatrix}. \quad (11)$$

This property is essential for testing models containing inequality constraints. Note that \mathcal{B} must be nonempty which implies that the models under investigation must be *comparable* (Mulder *et al.* 2010). For instance, $\mathcal{M}_1 : \beta = 0$ and $\mathcal{M}_2 : \beta > 1$ are not comparable but $\mathcal{M}_1 : \beta = 0$ and $\mathcal{M}_3 : \beta > 0$ are comparable. In this case, the CECPP of β is centered at $\beta = 0$. Note that $\mathcal{B} = \{0\}$ in this case. In the interface of **BIEMS** (Appendix A), it is automatically checked whether the constrained models are comparable when using the default prior specification setting.

3. The parameters that are compared with each other in the constrained models have equal variances in the CECPP. This is essential for incorporating the complexity of inequality constrained models with unequal weights on the parameters of interest, e.g., $\mathcal{M}_1 : \beta_1 > 2\beta_2 > 0$. This will be elaborated in Section 3.3.

In the remaining part of this section, it will be discussed how the CECPP and the corresponding posterior distribution is obtained.

3.2. Linear restrictions for constrained posterior priors

The CECPP can be seen as an average over all constrained posterior priors (CPPs). In contrast to standard posterior priors which are obtained by updating a noninformative improper prior with a minimal training sample, CPPs are obtained by updating a noninformative improper prior with a minimal training sample under certain linear restrictions. The linear restrictions ensure that the constrained posterior priors are located on the boundary of the constrained parameter space. Note that standard posterior priors are located around the likelihood, and therefore, generally will not satisfy the second property in Section 3.1. Mulder *et al.* (2010) showed that the resulting intrinsic Bayes factors may not be effective for testing inequality constrained models. In this section, it will be discussed how the linear restrictions are specified to obtain constrained posterior priors that satisfy the second property.

Two sets of linear restrictions will be constructed: One for the CPP means, say $[\mathbf{R}_{0,M} | \mathbf{r}_{0,M}]$, and one for the CPP variances, say $[\mathbf{R}_{0,V} | \mathbf{r}_{0,V}]$. The reason that different sets of restrictions are needed can for instance be seen when testing $\mathcal{M}_1 : \beta = 0$ versus $\mathcal{M}_0 : \beta \in R^1$. The

restriction for the CPP mean will be $\beta = 0$ to ensure that the CPP mean of β equals 0 (to satisfy the second property), while for the CPP variance of β no restriction is needed. Note that if $\beta = 0$ would be a restriction for the CPP variance, the CPP variance of β would become equal to zero which is undesirable. On the hand, when testing $\mathcal{M}_2 : \beta_1 > \beta_2 > \beta_3$ versus $\mathcal{M}_0 : \beta \in R^3$, the restriction for the CPP means and CPP variances are identical, namely,

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \beta = \mathbf{0}, \quad (12)$$

so that the CPP means are identical and the CPP variances are identical for β_1 , β_2 , and β_3 , for every randomly selected minimal training sample.

The restrictions for the CPP means, $[\mathbf{R}_{0,M}|\mathbf{r}_{0,M}]$, are identical to the row echelon form of $[\tilde{\mathbf{R}}_0|\tilde{\mathbf{r}}_0]$ where the zero rows are omitted. This ensures that the CPP means are located on the boundary \mathcal{B} . The use of the row echelon form with zero rows omitted is important for obtaining the hyperparameter of CPPs which is discussed in the following section.

The restrictions for the CPP variances are obtained from the row echelon form of $[\mathbf{R}_0^*|\mathbf{0}]$ where the zero rows are removed. Hence, the vector of constants does not play a role when restricting the CPP variances, i.e., $\mathbf{r}_{0,V} = \mathbf{0}$. The matrix \mathbf{R}_0^* is obtained from $\tilde{\mathbf{R}}_0$ as follows.

1. Rows of $\tilde{\mathbf{R}}_0$ that are permutations of $(w_{h,1}, 0, \dots, 0)$, with nonzero constants $w_{h,1} \in R^1$, are removed in \mathbf{R}_0^* . This is important for avoiding CPP variances of zero.
2. Rows of $\tilde{\mathbf{R}}_0$ that are permutations of $(w_{h,1}, w_{h,2}, 0, \dots, 0)$, with nonzero constants $w_{h,1}, w_{h,2} \in R^1$ where h is the row index, are replaced by the same permutations of the form $(\text{sign}(w_{h,1}), \text{sign}(w_{h,2}), 0, \dots, 0)$. These restrictions ensure that the CPP variances of β 's that are compared with each other have equal variances.
3. Rows of $\tilde{\mathbf{R}}_0$ that are permutations of $(w_{h,1}, \dots, w_{h,V}, 0, \dots, 0)$ with $V > 2$, i.e., rows that contain more than two nonzero elements, are replaced by a set of $V - 1$ rows that ensure that the corresponding β 's have equal variances. For instance, $\mathcal{M}_1 : \beta_2 - \beta_1 < \beta_3 - \beta_2$ results in a CPP mean restriction of $[1 \ -2 \ 1|0]$ and CPP variance restrictions given by (12).

If time-varying covariates are also incorporated in the model, additional restrictions of the form $\gamma^* = \mathbf{0}$ are automatically added to the mean restrictions $[\mathbf{R}_{0,M}|\mathbf{r}_{0,M}]$ and variance restrictions $[\mathbf{R}_{0,V}|\mathbf{0}]$. The mean and variance restrictions that were used to obtain the CPPs can be found at “View/Edit Default Prior” in the interface. Via “Specify restrictions manually”, a user can also specify the linear restrictions for β manually. However, we recommend to use the default setting as specified above.

In the stand-alone version of **BIEMS**, the restriction matrix for the CPP means can be manually specified in the file `restriction_matrix.txt`. The restrictions for the CPP variances are based on the rules stated above. By default, the system of restrictions is equal to the row echelon form of the system of constraints of the model \mathcal{M}_t , i.e., $[\tilde{\mathbf{R}}_t|\tilde{\mathbf{r}}_t]$. Note that the same restrictions must be used when computing B_{t0} . This is automatically done when using the **BIEMS** interface.

3.3. Hyper-parameters of the constrained posterior priors

Based on the linear restrictions for the means and the linear restrictions for the variances, CPPs are obtained from randomly selected minimal training samples, $[\mathbf{Y}_l|\mathbf{X}_l]$ where l is the training sample index, using restricted regression (Judge *et al.* 1980). The CPP of $\boldsymbol{\beta}$ is multivariate normally distributed,

$$\pi_0^C(\boldsymbol{\beta}|\mathbf{Y}_l) = \mathcal{N}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_l^R, \hat{\boldsymbol{\Psi}}_l^R), \quad (13)$$

where $\hat{\boldsymbol{\beta}}_l^R$ is the restricted maximum likelihood (RML) estimate that is obtained from the minimal training sample and the restrictions $[\mathbf{R}_{0,M}|\mathbf{r}_{0,M}]$ and $\hat{\boldsymbol{\Psi}}_l^R$ is the diagonal covariance matrix that is obtained from the minimal training sample and the linear restrictions $[\mathbf{R}_{0,V}|\mathbf{0}]$. The RML estimate of $\boldsymbol{\beta}$ maximizes the likelihood of $[\mathbf{Y}_l|\mathbf{X}_l]$ under the system of restrictions $\mathbf{R}_{0,M}\boldsymbol{\beta} = \mathbf{r}_{0,M}$, which yields

$$\hat{\boldsymbol{\beta}}_l^R = \hat{\boldsymbol{\beta}}_l + (\tilde{\mathbf{X}}_l^\top \tilde{\mathbf{X}}_l)^{-1} \mathbf{R}_{0,M}^\top [\mathbf{R}_{0,M} (\tilde{\mathbf{X}}_l^\top \tilde{\mathbf{X}}_l)^{-1} \mathbf{R}_{0,M}^\top]^{-1} (\mathbf{r}_{0,M} - \mathbf{R}_{0,M} \hat{\boldsymbol{\beta}}_l), \quad (14)$$

where $\tilde{\mathbf{X}}_l = \mathbf{I}_P \otimes \mathbf{X}_l$ and $\hat{\boldsymbol{\beta}}_l$ is the unrestricted ML estimate of $\boldsymbol{\beta}$ given $[\mathbf{Y}_l|\mathbf{X}_l]$. This is explained in more detail in Mulder *et al.* (2010). Because $\mathbf{R}_{0,M} \hat{\boldsymbol{\beta}}_l^R = \mathbf{r}_{0,M}$ holds, for all training samples, all CPPs are centered on the boundary under investigation \mathcal{B} .

Based on the minimal training sample and the restrictions $\mathbf{R}_{0,V}\boldsymbol{\beta} = \mathbf{0}$, restricted regression is used to obtain the covariance matrix, which is given by

$$\hat{\boldsymbol{\Psi}}_l^R = \frac{1}{N_l^*} \tilde{\boldsymbol{\Psi}}_l^R \left[\left(\boldsymbol{\Phi}_l^* [\tilde{\mathbf{X}}_l^{*\top} \tilde{\mathbf{X}}_l^*]^{-1} (\boldsymbol{\Phi}_l^*)^\top \right)_{diag} \right]^{-1}, \quad (15)$$

with

$$\begin{aligned} \tilde{\boldsymbol{\Psi}}_l^R &= \left[\boldsymbol{\Phi}_l (\tilde{\mathbf{X}}_l^\top \tilde{\mathbf{X}}_l)^{-1} \tilde{\mathbf{X}}_l^\top \left\{ [\hat{\boldsymbol{\Sigma}}_l^R]_{diag} \otimes \mathbf{I}_{N_l} \right\} \tilde{\mathbf{X}}_l (\tilde{\mathbf{X}}_l^\top \tilde{\mathbf{X}}_l)^{-1} \boldsymbol{\Phi}_l^\top \right]_{diag} \\ \hat{\boldsymbol{\Sigma}}_l^R &= \frac{1}{N_l} \left(\mathbf{Y}_l - \mathbf{X}_l \hat{\mathbf{B}}_l^R \right)^\top \left(\mathbf{Y}_l - \mathbf{X}_l \hat{\mathbf{B}}_l^R \right) \\ \boldsymbol{\Phi}_l &= \mathbf{I}_D - (\tilde{\mathbf{X}}_l^\top \tilde{\mathbf{X}}_l)^{-1} \mathbf{R}_{0,V}^\top [\mathbf{R}_{0,V} (\tilde{\mathbf{X}}_l^\top \tilde{\mathbf{X}}_l)^{-1} \mathbf{R}_{0,V}^\top]^{-1} \mathbf{R}_{0,V} \\ \boldsymbol{\Phi}_l^* &= \mathbf{I}_D - [\tilde{\mathbf{X}}_l^{*\top} \tilde{\mathbf{X}}_l^*]^{-1} \mathbf{R}_{0,V}^\top [\mathbf{R}_{0,V} [\tilde{\mathbf{X}}_l^{*\top} \tilde{\mathbf{X}}_l^*]^{-1} \mathbf{R}_{0,V}^\top]^{-1} \mathbf{R}_{0,V} \\ \tilde{\mathbf{X}}_l &= \mathbf{I}_P \otimes \mathbf{X}_l \\ \tilde{\mathbf{X}}_l^* &= \mathbf{I}_P \otimes \begin{bmatrix} \mathbf{d}_1^\top & \mathbf{1}_{K+PU}^\top \\ \vdots & \vdots \\ \mathbf{d}_N^\top & \mathbf{1}_{K+PU}^\top \end{bmatrix}, \end{aligned} \quad (16)$$

where $\mathbf{1}_{K+PU}$ is a vector of length $K + PU$ with ones, \mathbf{I}_D is an identity matrix of size $D = P(J + K + PU)$, $\hat{\boldsymbol{\Sigma}}_l^R$ is the RML estimate of the error covariance matrix $\boldsymbol{\Sigma}$, and $\text{vec}(\hat{\mathbf{B}}_l^R) = \hat{\boldsymbol{\beta}}_l^R$. See Mulder *et al.* (2010) for a thorough explanation of (15).

In (15), $\tilde{\boldsymbol{\Psi}}_l^R$ is the covariance matrix that is obtained when using restricted regression under the assumption of a diagonal covariance matrix $\boldsymbol{\Sigma}$ and independent variables in $\boldsymbol{\beta}$ in the constrained posterior prior. Because the training sample $[\mathbf{Y}_l|\mathbf{X}_l]$ of size $N_l = J + K + PU$

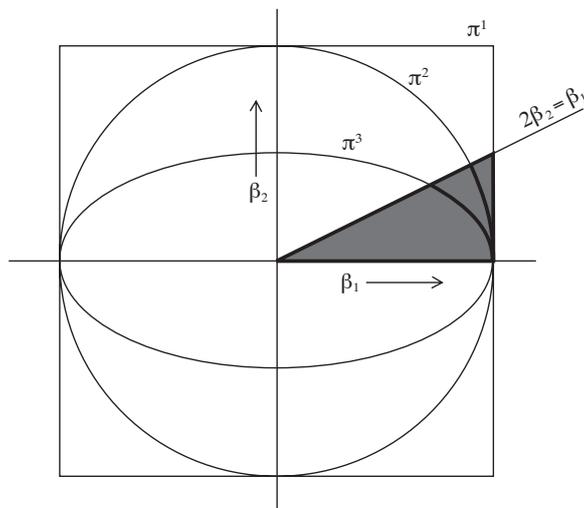


Figure 1: Contours of different priors: π^1 represents a uniform prior for (β_1, β_2) , π^2 represents identical normal priors for (β_1, β_2) , and π^3 represents normal priors with $\text{var}(\beta_1) = 4\text{var}(\beta_2)$. The complexity of the inequality constrained model $\mathcal{M}_1 : \beta_1 > 2\beta_2 > 0$ corresponds to the relative size of the grey areas. Hence, $c_1^1 = 0.0625$, $c_1^2 = 0.0738$, and $c_1^3 = 0.125$.

is minimal under the unrestricted model, and not under the restricted model, the variances are rescaled to a single observation by multiplying $\tilde{\Psi}_l^R$ with $\left[\left(\Phi_l^* [\tilde{\mathbf{X}}_l^{*\top} \tilde{\mathbf{X}}_l^*]^{-1} (\Phi_l^*)^\top \right)_{diag} \right]^{-1}$. Subsequently, when multiplying this with $1/N_l^*$, the variances are rescaled according to a training sample size of size N_l^* .

By default, the value of N_l^* is chosen equal to the minimal training sample size under the restricted model, which is equal to the number of free parameters under the restricted model. This is equal to the number of free parameters in the unrestricted model D minus the number of pivots in $[\mathbf{R}_{0,M} | \mathbf{r}_{0,M}]$, plus the number of diagonal elements of Σ , which equals P . The user can adjust the scale N_l^* to construct more or less informative constrained posterior priors. This can be changed in the interface under “Scale of prior variance”.

In the CPP covariance matrix $\hat{\Psi}_l^R$, the variances of parameters are equal which are compared with each other in the constrained models under investigation, i.e., $\text{var}(\beta_{d_1}) = \text{var}(\beta_{d_2})$ under the restriction $v_1\beta_{d_1} = v_2\beta_{d_2}$ for $\forall v_1, v_2 \neq 0$. Note that this differs from the CPP variance proposed by Mulder *et al.* (2010) where $v_1^2\text{var}(\beta_{d_1}) = v_2^2\text{var}(\beta_{d_2})$ under the same restriction. To justify this choice, we consider an example of an inequality constrained model with non-equal weights on the parameters of interest, e.g., $\mathcal{M}_1 : \beta_1 > 2\beta_2 > 0$. In Mulder *et al.* (2010), the complexity of an inequality constrained model was defined as the prior probability mass in the inequality constrained space assuming a uniform prior on $(-\ell, \ell)$ for $\ell \rightarrow \infty$. For model \mathcal{M}_1 , this implies that a measure of complexity that is equal to $1/16 = 0.0625$ according to this definition. In the Bayes factor based on CPPs, model complexity of \mathcal{M}_1 is incorporated as the CPP probability mass in the inequality constrained space $\{\boldsymbol{\beta} | \beta_1 > 2\beta_2 > 0\}$. Under the linear restrictions $\beta_1 = \beta_2 = 0$ for the CPP means and $\beta_1 = \beta_2$ for the CPP variances, such that $\text{var}(\beta_1) = \text{var}(\beta_2)$ and $E(\beta_1) = E(\beta_2) = 0$ in the CPP, model complexity is incorporated as approximately 0.0738. The probabilities 0.0625 and 0.0738 correspond to an appropriate measure of complexity, based on, respectively, identical uniform priors or identical normal

priors for β_1 and β_2 when evaluating $\mathcal{M}_1 : \beta_1 > 2\beta_2 > 0$. This is not the case for the variance restriction $\beta_1 = 2\beta_2$ (suggested by Mulder *et al.* 2010) yielding $\text{var}(\beta_1) = 4\text{var}(\beta_2)$ and a substantial different measure for the complexity for $\mathcal{M}_1 : \beta_1 > 2\beta_2 > 0$, namely 0.125. This is illustrated in Figure 1.

Note that when time-varying covariates are also incorporated in the model, the restrictions $\gamma^* = \mathbf{0}$, which fix the dummy parameters to zero, are also present in $[\mathbf{R}_{0,V}|\mathbf{0}]$. As a result, the variances of the parameters γ^* are equal to zero in $\hat{\Psi}_I^R$. Therefore, the CECPP covariance matrix of the parameters β^* , i.e., the full parameter vector β with the dummy parameters γ^* omitted, can be obtained by removing the rows and columns in $\hat{\Psi}_I^R$ that belong to the dummy parameters γ^* .

3.4. CECPP and posterior

The CECPP of β , given by

$$\pi_0^{EC}(\beta) = \mathcal{N}_{\beta}(\beta_0^{EC}, \Psi_0^{EC}), \quad (17)$$

is the conjugate form of the empirical expected-posterior prior (Pérez and Berger 2002) when using constrained posterior priors instead of standard posterior priors. Note here that Jeffreys' prior is used for the covariance matrix Σ . The hyper parameters of the CECPP of β are robustly estimated from 1,000,000 draws based on 1,000 draws from the constrained posterior priors of 1,000 randomly selected training samples according to the scheme in Table 1.

In step (vi), the CECPP means were calculated as the sample means of the RML estimates $\hat{\beta}_I^R$ when $\mathbf{R}_{0,M}$ contains rows with more than two non-zero elements. The reason is that, although the medians are more robust to outliers, the vector of medians ($\text{median}_I \beta_{I,1}^R, \dots, \text{median}_I \beta_{I,D}^R$)^T will generally not satisfy the mean restrictions $[\mathbf{R}_{0,M}|\mathbf{r}_{0,M}]$ in this case. Consequently, the second key property (Section 3.1) would not be satisfied when using the medians. In all other cases, the CECPP means are calculated as the sample medians. As a result, the CECPP will be located on the boundary of the constrained space of interest.

Due to sampling fluctuations and restrictions on the mean with unequal weights, e.g., $\beta_1 = 2\beta_2$, an additional modification of the CECPP standard deviations was performed at the end of step (vii) which was based on restricted regression as in (14). This ensures that the parameters that are compared with each other have equal CECPP variances. Consequently, the CECPP variances will satisfy the third key property (Section 3.1).

The conditional posterior distributions of $\beta|\Sigma, \mathbf{Y}$ and $\Sigma|\beta, \mathbf{Y}$ can be determined using Bayes' law from the CECPP and the likelihood in (5),

$$\begin{cases} \pi_0(\beta|\Sigma, \mathbf{Y}) = \mathcal{N}_{\beta}(\beta_N, \Psi_N) \\ \pi_0(\Sigma|\beta, \mathbf{Y}) = \mathcal{W}_{\Sigma}^{-1}(\Lambda_N, \nu_N) \end{cases} \quad (18)$$

with

$$\begin{aligned} \beta_N &= [(\Psi_0^{EC})^{-1} + \Sigma^{-1} \otimes \mathbf{X}^T \mathbf{X}]^{-1} [(\Psi_0^{EC})^{-1} \beta_0^{EC} + (\Sigma^{-1} \otimes \mathbf{X}^T \mathbf{X})\hat{\beta}] \\ \Psi_N &= [(\Psi_0^{EC})^{-1} + \Sigma^{-1} \otimes \mathbf{X}^T \mathbf{X}]^{-1} \\ \Lambda_N &= (\mathbf{Y} - \mathbf{X}\hat{\beta})^T (\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ \nu_N &= N + P + 1. \end{aligned}$$

-
- (i) Initialize the training sample size to $N_l = J + K + PU + 1$.
 - (ii) Randomly draw a minimal training sample $[\mathbf{Y}_l|\mathbf{X}_l]$ of size N_l .
 - (iii) Determine the hyperparameters $\{\hat{\boldsymbol{\beta}}_l^R, \hat{\boldsymbol{\Psi}}_l^R\}$ of the l th constrained posterior prior $\pi_0^C(\boldsymbol{\beta}|\mathbf{Y}_l)$.
 - (iv) Sample 1000 draws $\boldsymbol{\beta}_{0,l}^{(s)}$ from the constrained posterior prior of $\boldsymbol{\beta}$, for $s = 1000(l-1) + 1, \dots, 1000l$.
 - (v) Repeat steps (ii) to (iv) for $l = 1, \dots, 1000$.
 - (vi) The CECPP mean of $\boldsymbol{\beta}$ is calculated as $\boldsymbol{\beta}_0^{EC} = L^{-1} \sum_l \hat{\boldsymbol{\beta}}_l^R$ if $\mathbf{R}_{0,M}$ contains rows with more than 2 non-zero elements, or as $\beta_{0,d}^{EC} = \text{median}_l \hat{\beta}_{l,d}^R$, for $d = 1, \dots, D$, elsewhere.
 - (vii) Robust estimates (Huber 1981) are used for the CECPP covariance matrix of $\boldsymbol{\beta}$:

$$\tilde{\boldsymbol{\Psi}}_0^{EC}[d_1, d_2] = \frac{\hat{\sigma}^2(\beta_{d_1} + \beta_{d_2}) - \hat{\sigma}^2(\beta_{d_1} - \beta_{d_2})}{\hat{\sigma}^2(\beta_{d_1} + \beta_{d_2}) + \hat{\sigma}^2(\beta_{d_1} - \beta_{d_2})} \hat{\sigma}(\beta_{d_1}) \hat{\sigma}(\beta_{d_2}), \text{ for } d_1, d_2 = 1, \dots, D$$

where $\hat{\sigma}(\beta_d) = \text{median}_{s,l}(|\beta_{0,l,d}^{(s)} - \beta_{0,d}^{EC}|) / \Phi^{-1}(.75)$, for $s = 1, \dots, 1000$
with $\Phi^{-1}(\cdot)$ the inverse of the cdf of the standard normal distribution

The standard deviations and the correlation matrix of $\tilde{\boldsymbol{\Psi}}_0^{EC}$ are separated according to $\tilde{\boldsymbol{\Psi}}_0^{EC} = \text{diag}(\tilde{\mathbf{s}}_0^{EC}) \mathbf{C}_0^{EC} \text{diag}(\tilde{\mathbf{s}}_0^{EC})$. Subsequently, $\boldsymbol{\Psi}_0^{EC} = \text{diag}(\mathbf{s}_0^{EC}) \mathbf{C}_0^{EC} \text{diag}(\mathbf{s}_0^{EC})$ with $\mathbf{s}_0^{EC} = (\mathbf{I}_D - \mathbf{R}_{0,V}^\top (\mathbf{R}_{0,V} \mathbf{R}_{0,V}^\top)^{-1} \mathbf{R}_{0,V}) \tilde{\mathbf{s}}_0^{EC}$.

Table 1: CECPP construction scheme.

3.5. Issues with minimal training samples

Because the β 's are modeled independently in the constrained posterior prior ($\hat{\boldsymbol{\Psi}}_l^R$ is diagonal), training samples of size $J + K + PU + 1$ generally result in appropriate constrained posterior priors. However, in some cases, e.g., data with many identical observations, the diagonal elements of these matrices may be (approximately) zero. This problem can be overcome by increasing the training sample size such that the variance in the training data increases. In **BIEMS**, the training sample size is increased by 1 after 5 training samples have been sequentially selected for which the estimate of the generalized error variance based on the training sample, i.e., $\det(\hat{\boldsymbol{\Sigma}}_l^R/N_l)$, is smaller than 0.05 times the estimate of the generalized error variance of the complete data matrix, i.e., $\det(\mathbf{S}/N)$, where $\hat{\boldsymbol{\Sigma}}_l^R$ is the estimate of $\boldsymbol{\Sigma}$ based on the RML estimate of \mathbf{B} of the l th training sample given in (16). Therefore, the scheme in Table 2 is used where the roman numbers refer to the steps in Table 1. Note that the constrained posterior prior variance of $\boldsymbol{\beta}$ is not influenced by the size of the minimal training sample because the constrained posterior prior variances of the β 's are scaled according to the number of free parameters under the restriction $[\mathbf{R}_{0,M}|\mathbf{r}_{0,M}]$, which remains unchanged.

-
1. Initialize $N_l = J + K + PU$, and calculate $\det(\mathbf{S}/N)$ with $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})^\top(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$.
 2. Set $N_l = N_l + 1$, $V = 0$, and the training sample counter to 1.
 3. Do steps (ii), and (iii), and determine $\det(\hat{\Sigma}_l^R)$.
 4. If $\det(\hat{\Sigma}_l^R/N_l) < 0.05 \cdot \det(\mathbf{S}/N)$, then “bad training sample”, set $V = V + 1$, go to 5. Else, go to 6.
 5. If $V < 5$, go to 3. Else, discard all constrained posterior prior draws and go to 2.
 6. Set $V = 0$. Do step (v). If $l < 1,000$, go to 2, else, do step (vii).
-

Table 2: Training sample size adjustment scheme.

4. Calculating Bayes factors

When the CECPP is used as prior under the unconstrained model \mathcal{M}_0 and the prior under the inequality constrained model \mathcal{M}_t is equal to the CECPP truncated in the parameter space of the constrained model \mathcal{M}_t , denoted by M_t , the Bayes factor of model \mathcal{M}_t against model \mathcal{M}_0 can be expressed by

$$\begin{aligned} B_{t0} &= \frac{f_t}{c_t} \\ &= \frac{\int_{\beta \in \mathcal{M}_t} \pi_0(\beta|\mathbf{Y}) \partial\beta}{\int_{\beta \in \mathcal{M}_t} \pi_0^{EC}(\beta) \partial\beta}, \end{aligned} \quad (19)$$

where $\pi_0(\beta|\mathbf{Y})$ is the marginal posterior of β that is based on the CECPP and the complete data \mathbf{Y} . The derivation can be found in [Klugkist and Hoijtink \(2007\)](#). They refer to the unconstrained prior as the encompassing prior.

BIEMS automatically computes all Bayes factors B_{t0} of each constrained model \mathcal{M}_t against the unconstrained model \mathcal{M}_0 . Because the same CECPP is used for calculating the Bayes factors B_{t0} , for $t = 1, \dots, T$, the Bayes factor between two constrained models \mathcal{M}_1 and \mathcal{M}_2 can be computed by

$$B_{12} = \frac{B_{10}}{B_{20}}.$$

For an inequality constrained model \mathcal{M}_t , **BIEMS** automatically computes the Bayes factor of model \mathcal{M}_t versus its complement, i.e., $\mathcal{M}_{t'} : \beta \notin M_t$, which equals

$$B_{tt'} = \frac{f_t}{1 - f_t} \frac{1 - c_t}{c_t}.$$

In the **BIEMS** interface ([Appendix A](#)), the posterior model probabilities can be computed by pressing the “Compute” button in the “Posterior model probabilities” frame. These are calculated according to

$$P(\mathcal{M}_t|\mathbf{Y}) = \frac{B_{t0}}{\sum_{t=0}^T B_{t0}}. \quad (20)$$

4.1. Estimating f_t and c_t for an inequality constrained model

The quantities f_t and c_t are estimated using a Gibbs sampler (e.g., Gelman *et al.* 2004). The complexity c_t can be estimated by sampling S draws from the multivariate normal CECPP of β in (17). The estimate is equal to the proportion of draws satisfying the inequality constraints of model \mathcal{M}_t , i.e.,

$$\hat{c}_t = S^{-1}I(\beta_0^{(s)} \in M_t), \quad (21)$$

where $\beta_0^{(s)}$ is the s th draw from the CECPP, for $s = 1, \dots, S$ ($S = 20,000$ by default, modifiable in the interface, see Appendix A, or in the file `input_BIEMS.txt`, see Appendix B), and $I(\cdot)$ is the identity function. Similarly, f_t can be estimated as the proportion of draws from the unconstrained posterior in (18) that satisfy the inequality constraints of model \mathcal{M}_t , i.e.,

$$\hat{f}_t = S^{-1}I(\beta_N^{(s)} \in M_t), \quad (22)$$

where $\beta_N^{(s)}$ is the s th draw from the unconstrained posterior. Based on the posterior conditionals of $\beta|\Sigma, \mathbf{Y}$ and $\Sigma|\beta, \mathbf{Y}$ in (18), a sample from the unconstrained posterior is obtained using the MCMC procedure in Table 3.

In the first step, the initial value of Σ is obtained using the posterior estimate $\hat{\mathbf{B}}_N$, with $\text{vec}(\hat{\mathbf{B}}_N) = \hat{\beta}_N = [(\Psi_0^{EC})^{-1} + \hat{\Sigma}^{-1} \otimes \mathbf{X}^\top \mathbf{X}]^{-1} [(\Psi_0^{EC})^{-1} \beta_0^{EC} + (\hat{\Sigma}^{-1} \otimes \mathbf{X}^\top \mathbf{X})\hat{\beta}]$, which is a weighted average of the prior mean of \mathbf{B}_0^{EC} (with $\text{vec}(\mathbf{B}_0^{EC}) = \beta_0^{EC}$) and the ML estimate $\hat{\mathbf{B}}$ (with $\text{vec}(\hat{\mathbf{B}}) = \hat{\beta}$).

More efficient estimation for small c_t

Estimating c_t and f_t using the above described method can be computationally intensive for models with small values for these probabilities. For example, for $c_1 = 1/10! \approx 2.76e-7$ corresponding to a model $\mathcal{M}_1 : \beta_1 > \dots > \beta_{10}$ based on the CECPP with identical priors for each β . A sample of at least $1e9$ draws is needed to obtain a fair estimate of c_1 . For this reason, c_t and f_t are estimated in steps of two inequalities. For model \mathcal{M}_1 , the probabilities

-
1. Set the initial value $\Sigma^{(0)} = N^{-1}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}_N)^\top (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}_N)$.
 2. Draw $\beta_N^{(s)}$ from $\pi_0(\beta|\Sigma^{(s-1)}, \mathbf{Y})$ in (18).
 3. Draw $\Sigma^{(s)}$ from $\pi_0(\Sigma|\beta^{(s)}, \mathbf{Y})$ in (18).
 4. Repeat steps 2 and 3 for $s = 1, \dots, S$.
-

Table 3: MCMC sampling from the posterior scheme.

c_t and f_t can be expressed stepwise as

$$\begin{aligned} P(\beta_1 < \dots < \beta_{10}) &= P(\beta_1 < \beta_2 < \beta_3)P(\beta_3 < \beta_4 < \beta_5|\beta_1 < \beta_2 < \beta_3) \cdot \\ &P(\beta_5 < \beta_6 < \beta_7|\beta_1 < \dots < \beta_5)P(\beta_7 < \beta_8 < \beta_9|\beta_1 < \dots < \beta_7) \\ &P(\beta_9 < \beta_{10}|\beta_1 < \dots < \beta_9) \\ &= c_{1,1} \cdot \dots \cdot c_{1,5} \text{ for the prior} \end{aligned} \tag{23}$$

$$= f_{1,1} \cdot \dots \cdot f_{1,5} \text{ for the posterior,} \tag{24}$$

where the second index of c and f denotes the step.

Based on the CECPP, these (conditional) probabilities are equal to $\frac{1}{6}$, $\frac{1}{20}$, $\frac{1}{42}$, $\frac{1}{72}$, and $\frac{1}{10}$, respectively. The first probability $P(\beta_1 < \beta_2 < \beta_3)$ can be estimated by directly sampling from the unconstrained CECPP and posterior and the proportion of draws satisfying $\beta_1 < \beta_2 < \beta_3$ are the estimates c_1 and f_1 , respectively. The estimates of the conditional probabilities are equal to the proportion of draws satisfying each pair of inequality constraints based on a sample from the conditional distributions of $\beta_d|\beta_{-d}, \Sigma$, where β_{-d} is the vector of β 's with β_d omitted. If β_d is constrained by β_{-d} , the conditional distribution is truncated normal, i.e., $\beta_d|\beta_{-d}, \Sigma \sim \mathcal{N}(m_{\beta_d}, s_{\beta_d}^2; \beta_L, \beta_U)$ where m_{β_d} and $s_{\beta_d}^2$ are the conditional mean and variance of β_d , respectively, and β_L and β_U are the lower and upper bound of β_d , respectively. If β_d is unconstrained, i.e., $\beta_d|\beta_{-d}, \Sigma \sim \mathcal{N}(m_{\beta_d}, s_{\beta_d}^2)$, the conditional distribution is normal. Sampling from the truncated normal distribution can be done using inverse probability sampling (Gelfand *et al.* 1992). The number of draws to estimate these conditional probabilities is 20,000 by default. Thus, the total number of draws to estimate c_1 is 100,000. The same methodology is used to calculate the posterior probability f_1 .

Burn-in and convergence

BIEMS does not provide explicit MCMC output because sampling the mean parameter vector β and the covariance matrix Σ is relatively straightforward in the multivariate normal linear model using the Gibbs sampler presented in Table 3. By starting at the posterior expectation of Σ , no burn-in period is needed for obtaining a posterior sample, and consequently, convergence is not an issue. The same holds for obtaining a sample from the CECPP of β which is multivariate normal and independent of Σ .

When sampling under inequality constraints as was discussed above, the starting values for β are equal to the ML estimates. When the starting values are not located in the allowed parameter space, e.g., the ML estimate equals $(\hat{\beta}_1, \dots, \hat{\beta}_5) = (1, 1, 1, 1, 0)$ when calculating $P(\beta_3 < \beta_4 < \beta_5|\beta_1 < \beta_2 < \beta_3)$, a burn-in period of 1,000 iterations is used to obtain a starting value for β and Σ in the highest probability density region under the restrictions at hand. The length of this burn-in period was always sufficient during the debugging of the program, also in extreme situations.

Furthermore, by default different seeds are used for every MCMC run which always resulted in accurate estimates of the Bayes factor during the debugging of the program. If necessary, the user can increase the MCMC sample size to obtain an estimate of the Bayes factor with higher accuracy. The computation of the sampling error of the Bayes factor is discussed later in this section.

Inverse probability sampling when the normal density is approximately zero

If the posterior density of β_p is approximately zero in the allowed parameter space, say (β_L, β_U) , inverse probability sampling (Gelfand *et al.* 1992), according to $\beta^{(s)} = \Phi^{-1}(u^{(s)})$, with $u^{(s)} \sim U(\Phi(\beta_L), \Phi(\beta_U))$ where Φ is the conditional posterior cdf of β , does not work because $\Phi(\beta_L) \approx \Phi(\beta_U) \approx 0$ or 1. This occurs for instance when the constraints badly fit the data.

When the allowed parameter space is an interval (β_L, β_U) with approximately zero density, which is the case when the conditional posterior is $\beta \sim N(m_\beta, s_\beta^2)$ with $m_\beta \ll \beta_L$ or $m_\beta \gg \beta_U$, the conditional posterior of β in the interval (β_L, β_U) is approximated using an exponential distribution, i.e., $\pi(\beta) = \lambda \exp(-\lambda\beta)$. If $m_\beta \ll \beta_L$, the parameter λ is obtained by first determining the distance ξ from the lower bound β_L for which the posterior density is halved, i.e.,

$$\frac{\phi(\beta_L + \xi; m_\beta, s_\beta^2)}{\phi(\beta_L; m_\beta, s_\beta^2)} = \frac{1}{2} \Rightarrow \xi = m_\beta - \beta_L + \sqrt{(\beta_L - m_\beta)^2 + 2 \log(2) s_\beta^2}.$$

Because the exponential density is halved at $\log(2)/\lambda$, it holds that $\lambda = \log(2)/\xi$ with ξ given above. Based on this λ , a draw is obtained from the exponential distribution, $\beta^{(e)} \sim \exp(\lambda; \beta_U)$ truncated in the interval $[0, \beta_U - \beta_L]$. Subsequently, $\beta^{(s)} = \beta_L + \beta^{(e)}$. In the case of $m_\beta \gg \beta_U$, the procedure is similar.

4.2. Sampling error of the estimate of the Bayes factor

The estimation of the probabilities f_t and c_t as proportions of, respectively, the posterior and prior draws that satisfy the inequality constraints results in an estimation error of the Bayes factor. This estimation error is proportional to the reciprocal of the sample size. In **BIEMS**, this estimation error is reported in the output file. In this section, it is illustrated how the standard error of the estimate is calculated.

Again, we consider the inequality constrained model $\mathcal{M}_1 : \mu_1 < \dots < \mu_{10}$. Based on (23) and (24), the Bayes factor of \mathcal{M}_1 versus \mathcal{M}_0 is estimated as

$$\hat{B}_{10} = \frac{\hat{f}_{1,1} \cdot \dots \cdot \hat{f}_{1,5}}{\hat{c}_{1,1} \cdot \dots \cdot \hat{c}_{1,5}}. \quad (25)$$

In **BIEMS**, the estimates of these probabilities are obtained as the proportion of hits in the CECPP and posterior, for $c_{1,h}$ and $f_{1,h}$, $h = 1, \dots, 5$, respectively, satisfying the inequality constraints. Hence, each probability has a beta distribution given by $beta(\ell_h - 1, S - \ell_h - 1)$, where ℓ_h is the number of draws satisfying the inequality constraints in step h from a total of S draws from CECPP or posterior. Subsequently, a sample of the Bayes factor B_{10} can be obtained by sampling from these beta-distributions. The standard error of the estimated Bayes factor B_{10} is equal to the standard deviation of this sample.

A small simulation study was conducted to determine the relationship between the sample size per two inequalities and the estimation error of the Bayes factor. A data set of size $N = 10$ was generated from a population with means $\beta = (0, 0, 0, 1, 1, 1, 2, 2, 2, 2)$ and covariance matrix $\Sigma = \mathbf{I}_{10}$. The Bayes factor of $\mathcal{M}_1 : \beta_1 < \dots < \beta_{10}$ versus $\mathcal{M}_0 : \beta \in R^{10}$ was equal to 2.0e6. In Figure 2, a graph is displayed where the standard error of the Bayes factor is plotted versus the sample size per two inequalities. Given the large outcome of the Bayes

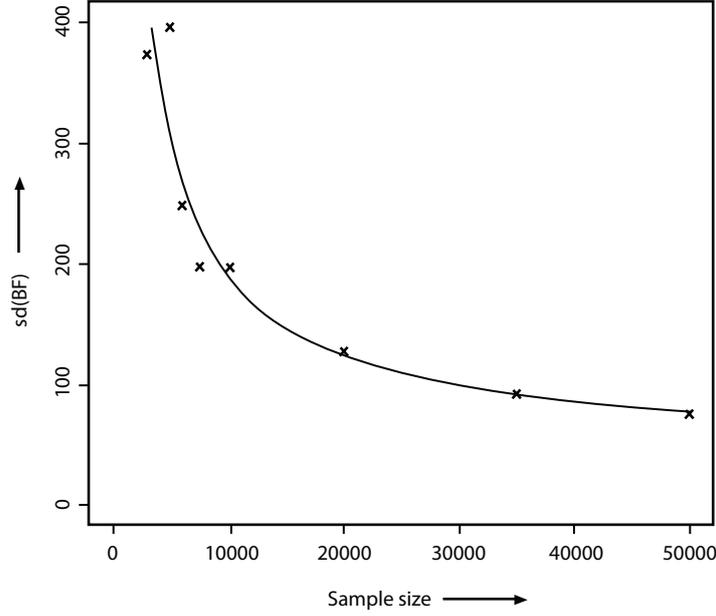


Figure 2: Regression line of the sample size per two inequality constraints versus the standard error of the Bayes factor. The dots are the actual measurements of the standard errors. The Bayes factor of this data set is equal to 2.0e6.

factor, these standard errors are reasonably low. The default sample size is equal to 20,000 iterations per two inequalities. The sample size can be adjusted in the interface (Appendix A) and in the file `input_BIEMS.txt` (Appendix B) if a more accurate estimate is needed.

4.3. Equality constrained models

For constrained models with equality constraints between the model parameters (possibly in addition to inequalities), f_t and c_t cannot be estimated by directly using the above described method because sampled values are never exactly equal. In this section, it is described how the Bayes factor of such models versus the unconstrained model is estimated in **BIEMS** based on the methodology given in the previous section. The idea was suggested by (Laudy 2006, p. 115).

First, the equality constraints in model \mathcal{M}_t are replaced by approximate equalities, e.g.,

$$\beta_1 = \beta_2 \Rightarrow \beta_1 \approx \beta_2 \Leftrightarrow \begin{cases} \beta_1 - \beta_2 > -\delta_0 \\ \beta_2 - \beta_1 > -\delta_0. \end{cases}$$

The evaluation bound δ_0 is chosen equal to the CECPP standard error of β_1 . Thus, the constrained model \mathcal{M}_t is approximated by an inequality constrained model, which is denoted by \mathcal{M}_{t,δ_0} . The Bayes factor of model \mathcal{M}_{t,δ_0} versus \mathcal{M}_0 , i.e., $B_{(t,\delta_0)0}$, can be estimated using the method of the previous section.

Subsequently, the Bayes factor is estimated between the inequality constrained model \mathcal{M}_{t,δ_1} and the inequality constrained model \mathcal{M}_{t,δ_0} , with $\delta_1 = \delta_0/2$, as the ratio of proportion of draws from the posterior and CECPP truncated at the inequality constraints of model \mathcal{M}_{t,δ_0} that satisfy the inequality constraints of model \mathcal{M}_{t,δ_1} . CECPP and posterior samples under

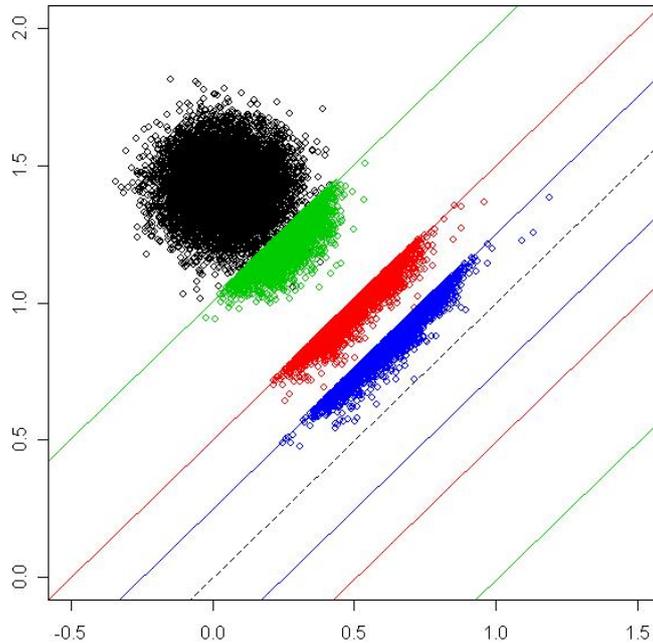


Figure 3: Draws from the unconstrained posterior of (β_1, β_2) (black), and subsequent constrained posteriors by $|\beta_1 - \beta_2| < \delta_w$, with $\delta_{a+1} = \delta_a/2$, $\delta_0 = 0.9$. The dashed line is $\beta_1 = \beta_2$. The data was generated for two groups of size 100 each with means $\beta = (0, 1.5)^\top$ and error variance $\sigma^2 = 1$.

the inequality constrained model $\mathcal{M}_{t, \delta_0}$ can be obtained using inverse probability sampling (Gelfand *et al.* 1992) and the method described in Section 4.1.

Then, the Bayes factor is estimated between the inequality constrained model $\mathcal{M}_{t, \delta_a}$ and the inequality constrained model $\mathcal{M}_{t, \delta_{a-1}}$, with $\delta_a = \delta_{a-1}/2$, for $a = 2, \dots, A$ until the Bayes factor is approximately one, i.e., $|\hat{B}_{(t, \delta_A)(t, \delta_{A-1})} - 1| < 0.01$. The Bayes factor of model \mathcal{M}_t versus \mathcal{M}_0 is then estimated as

$$\hat{B}_{t0} = \hat{B}_{(t, \delta_0)0} \prod_{a=1}^A \hat{B}_{(t, \delta_a)(t, \delta_{a-1})}.$$

As an example, data was generated of size 100 for each of two groups with means $\mu = (0, 1.5)^\top$ and covariance matrix $\Sigma = \mathbf{I}_2$. The model $\mathcal{M}_1 : \beta_1 = \beta_2$ was tested against $\mathcal{M}_0 : \beta \in R^2$. In Figure 3, a sample is displayed from the unconstrained posterior of (β_1, β_2) (black) and from the subsequent posteriors of (β_1, β_2) under the constraint $|\beta_1 - \beta_2| < \delta_a$, for $a = 0$ and 1. The resulting Bayes factor of $\mathcal{M}_1 : \beta_1 = \beta_2$ against the unconstrained model was equal to 0.

The Savage-Dickey density ratio

Consider the simple selection problem of an equality constrained model $\mathcal{M}_1 : \beta = m$ versus the unconstrained model $\mathcal{M}_0 : \beta \in R^1$. The Bayes factor in (19) is then equal to the Savage-Dickey density ratio (Dickey, 1971), which is given by

$$B_{10} = \frac{\pi_0(\beta = m | \mathbf{x})}{\pi_0(\beta = m)}. \quad (26)$$

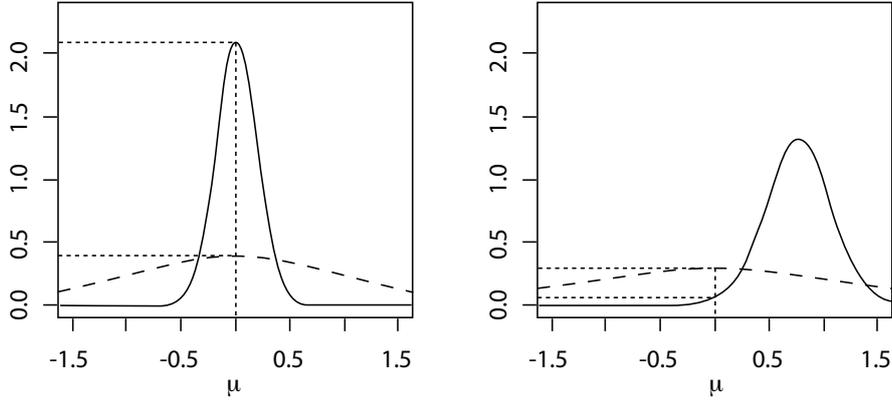


Figure 4: Prior (dashed line) and posterior (solid line) sample density plots of a data set of size $N = 30$ and $\mathcal{N}(0, 1)$ (left panel) and a data set of size $N = 15$ and $\mathcal{N}(.7, 1)$ (right panel). Using the Savage-Dickey density ratio, the Bayes factors of $\mathcal{M}_1 : \beta = 0$ versus $\mathcal{M}_0 : \mu \in \mathbb{R}^1$ equal $B_{10} \approx \frac{2.1}{0.4} = 5.25$ (left panel) and $B_{10} \approx \frac{0.07}{0.29} = 0.24$ (right panel).

Hence, the Bayes factor is given by the unconstrained posterior density of β evaluated at m divided by the unconstrained prior density of β evaluated at m . [Wetzels et al. \(2010\)](#) provide a proof that (19) converges to the Savage-Dickey density ratio for the selection problem $\mathcal{M}_{1,\delta} : |\beta| < \delta$ (for some value $\delta > 0$) versus $\mathcal{M}_0 : \beta \in \mathbb{R}^1$ as $\delta \downarrow 0$.

Two data sets were generated of size $N = 30$ with $\mathcal{N}(0, 1)$ and of size $N = 15$ with $\mathcal{N}(.7, 1)$. The sample density plots of the CECPP and the posterior are displayed in Figure 4. Furthermore, consider the model selection problem $\mathcal{M}_1 : \beta = 0$ versus $\mathcal{M}_0 : \mu \in \mathbb{R}^1$. From the figure, it can be concluded that the Bayes factor of model \mathcal{M}_1 versus \mathcal{M}_0 is equal to $B_{10} \approx \frac{2.1}{0.4} = 5.25$ (left panel) and $B_{10} \approx \frac{0.07}{0.29} = 0.24$ (right panel) using the Savage-Dickey density ratio in (26). The Bayes factors provided by **BIEMS** using the above described mechanism were equal to $B_{10} = 5.35$ (with sampling error $7.4e-2$) for the first data set and $B_{10} = 0.24$ (with sampling error $4.6e-3$). Hence, the true Bayes factors lie well within the confidence intervals provided by **BIEMS** from which it can be concluded that the methodology can also be used for models containing equality constraints between the parameters. The effectiveness of this procedure was also shown by [van Wesel et al. \(2011\)](#).

5. Short examples

In this section several commonly used models are briefly discussed including two constrained models of interest. In each example, the parameter matrix \mathbf{B} , the vector $\boldsymbol{\beta}$, and different constrained models are given. Also the specification of the restriction matrix is discussed in each example. In Section 5.6, five empirical studies are briefly discussed which were analyzed using **BIEMS**. The analyses of these studies can be found on the internet site <http://tinyurl.com/informativehypotheses>.

5.1. 2 by 3 ANOVA

Consider an ANOVA design with two factors $a = 1, 2$ and $b = 1, 2, 3$. The model can be

implemented in **BIEMS** as

$$\mathbf{B} = \boldsymbol{\beta} = (\mu_{11}, \mu_{21}, \mu_{12}, \mu_{22}, \mu_{13}, \mu_{23})^\top.$$

Note here that the order of the μ 's in \mathbf{B} and $\boldsymbol{\beta}$ depends on the coding of the groups 1 to 6 in `data.txt`. Subsequently, the following model selection problem tests the presence of an interaction effect in a specific direction

$$\mathcal{M}_1 : \mu_{11} - \mu_{21} > \mu_{12} - \mu_{22} > \mu_{13} - \mu_{23}$$

$$\mathcal{M}_2 : \mu_{11} - \mu_{21} = \mu_{12} - \mu_{22} = \mu_{13} - \mu_{23}.$$

The constraints of each model needs to be separately specified in the interface. For instance for \mathcal{M}_1 , the inequality constraints are $\mu_{11} - \mu_{21} > \mu_{12} - \mu_{22}$ and $\mu_{12} - \mu_{22} > \mu_{13} - \mu_{23}$. The set of model constraints is then

$$\mathcal{M}_1 : \tilde{\mathbf{R}}_1 \boldsymbol{\beta} >_t \tilde{\mathbf{r}}_1 \Rightarrow \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 \end{bmatrix} \boldsymbol{\beta} > \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

For \mathcal{M}_2 , the inequality “>” is replaced by an equality sign “=”. The default set of restrictions for the specification of the prior distribution are equal to the row echelon forms of

$$\mathbf{R}_{0,M} \boldsymbol{\beta} = \mathbf{r}_{0,M} \Rightarrow \begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for the CPP means and

$$\mathbf{R}_{0,V} \boldsymbol{\beta} = \mathbf{0} \Rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

for the CPP variances. The resulting CECPP means and variances also satisfy these restrictions.

5.2. Regression 1: Inequality constrained models

Consider a univariate regression model with five covariates, i.e.,

$$\mathbf{B} = \boldsymbol{\beta} = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)^\top,$$

and the following inequality constrained models

$$\mathcal{M}_1 : \alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 > \alpha_5$$

$$\mathcal{M}_2 : \alpha_1 > (\alpha_2, \alpha_3, \alpha_4, \alpha_5).$$

The (in)equality constraints in each model must be separately specified in the model, e.g., $\alpha_1 > \alpha_2$, $\alpha_1 > \alpha_3$, $\alpha_1 > \alpha_4$, and $\alpha_1 > \alpha_5$ for model \mathcal{M}_2 . The default set of restrictions are obtained by taking the row echelon form of the combination of the constraints of both models.

This results in an identical set of restrictions for the CPP means and variances which is given by

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix} \boldsymbol{\beta} = \mathbf{0}.$$

Based on these restrictions, the CECPP of $\alpha_1, \dots, \alpha_5$ are identical. After obtaining the Bayes factors B_{10} and B_{20} , the Bayes factor between these two models is equal to $B_{12} = B_{10}/B_{20}$. The posterior model probabilities can also be computed in the interface of **BIEMS** (Appendix A).

When the data is not standardized, the constraints of models \mathcal{M}_1 and \mathcal{M}_2 are evaluated for unstandardized regression coefficients. Generally this is not recommendable because covariates may be measured on different scales. For this reason, it is recommended to standardize the data which can be specified in **Step 2** in the interface (Appendix A) and in the file `input_biems.txt` (Appendix B). As a result, the model constraints are evaluated for standardized regression coefficients.

5.3. Regression 2: Variable selection

Again, consider the regression model above with five exploratory variables and the following model selection problem

$$\begin{aligned} \mathcal{M}_1 & : \alpha_2 = 0, \alpha_3 = 0, \alpha_4 = 0, \alpha_5 = 0 \\ \mathcal{M}_2 & : \alpha_3 = 0, \alpha_5 = 0. \end{aligned}$$

This is a variable selection problem where model \mathcal{M}_1 only incorporates the first covariate to explain the dependent variable, model \mathcal{M}_2 incorporates the first, second, and fourth covariate, and the unconstrained model \mathcal{M}_0 incorporates all five covariates. Similarly as in the previous section, we recommend to standardize the data.

In the interface where both models are simultaneously implemented, the default set of restrictions for the CPP means are given by

$$\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{\beta} = \mathbf{0}.$$

In the stand-alone version where each model is separately evaluated, this set of restrictions must be manually set for model \mathcal{M}_2 (for \mathcal{M}_1 , these restrictions correspond with the default setting). There are no restrictions for the CPP variances in the default setting, and therefore the CECPP variances will generally differ between the α 's.

5.4. MANCOVA

Consider an experiment where three outcome variables are measured for two different groups with one covariate, i.e.,

$$\mathbf{B} = \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \\ \alpha_{11} & \alpha_{12} & \alpha_{13} \end{bmatrix},$$

and $\boldsymbol{\beta} = (\mu_{11}, \mu_{21}, \alpha_{11}, \dots, \alpha_{13})^\top$, and the following models

$$\begin{aligned}\mathcal{M}_1 &: \mu_{11} > \mu_{21}, \mu_{12} < \mu_{22}, \mu_{13} > 2\mu_{23} > 0 \\ \mathcal{M}_2 &: \boldsymbol{\mu} \notin \mathcal{M}_1.\end{aligned}$$

Under model \mathcal{M}_1 , the third measurement mean of the first group μ_{13} is more than twice as large as the measurement mean of the second group μ_{23} , which is larger than zero. Such constraints can occur in practice when the μ 's are the means of measurement differences.

The default set of restrictions for the CPP means are equivalent with

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0. \end{bmatrix} \boldsymbol{\beta} = \mathbf{0},$$

and for the CPP variances, the restrictions yield

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0. \end{bmatrix} \boldsymbol{\beta} = \mathbf{0},$$

We also recommend to standardize the explanatory variables in a MANCOVA. The Bayes factor of model \mathcal{M}_1 against its complement \mathcal{M}_2 is automatically calculated, which can be viewed by pressing the button “Detailed output” after the Bayes factors have been calculated.

5.5. Repeated measurements

A model with five repeated measurements and one time-varying covariate is modeled as

$$\mathbf{B} = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 & \mu_5 \\ \gamma_{1,1} & 0 & 0 & 0 & 0 \\ 0 & \gamma_{1,2} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{1,3} & 0 & 0 \\ 0 & 0 & 0 & \gamma_{1,4} & 0 \\ 0 & 0 & 0 & 0 & \gamma_{1,5} \end{bmatrix},$$

and $\boldsymbol{\beta} = (\mu_1, \gamma_{1,1}, \mu_2, \gamma_{1,2}, \mu_3, \gamma_{1,3}, \mu_4, \gamma_{1,4}, \mu_5, \gamma_{1,5})^\top$. As an example, we consider two models with constraints on the measurement adjusted means $\boldsymbol{\mu}$ (constraints between the γ 's can be specified in a similar way)

$$\begin{aligned}\mathcal{M}_1 &: \mu_1 > \mu_2 > \mu_3 > \mu_4 > \mu_5 \\ \mathcal{M}_2 &: \mu_1 \approx \mu_2 \approx \mu_3 \approx \mu_4 \approx \mu_5 \text{ with } \delta = 0.5.\end{aligned}$$

In model \mathcal{M}_2 , the maximal difference between subsequent measurement means is $\delta = 0.5$, i.e., $|\mu_p - \mu_{p+1}| < 0.5 \Leftrightarrow \mu_p - \mu_{p+1} > -0.5, -\mu_p + \mu_{p+1} > -0.5$, for $p = 1, \dots, 4$. Therefore, 8 inequality constraints need to be specified for \mathcal{M}_2 , namely, $\mu_1 - \mu_2 > -0.5, \mu_2 - \mu_1 > -0.5, \dots, \mu_4 - \mu_5 > -0.5, \mu_5 - \mu_4 > -0.5$. It is important that pairs of inequality constraints

that form an approximately equality must be specified next to each other. This is essential in order to obtain proper default restrictions for the CPPs, which, in this case, are given by

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix} \boldsymbol{\beta} = \mathbf{0},$$

for both the CPP means and CPP variances.

The time-varying covariate needs to be standardized, each response $w_{i,1,p}$ is standardized based on the sample mean $\bar{w}_1 = (NP)^{-1} \sum_{i,p} w_{i1p}$ and sample standard deviation $s_{w_1} = \sqrt{(NP)^{-1} \sum_{i,p} (w_{i1p} - \bar{w}_1)^2}$ of the time-varying covariate.

5.6. Empirical studies

Here, we briefly discuss five empirical studies which were analyzed using **BIEMS**. Replication materials are available along with the manuscript and the data sets can also be found on the website <http://tinyurl.com/informativehypotheses>. The first data set was discussed by [Kammers *et al.* \(2009\)](#), the second data was discussed by [Mulder *et al.* \(2009\)](#), and the third, fourth, and fifth data sets were discussed by [Mulder *et al.* \(2010\)](#).

Repeated measurements 1: Rubber hand illusion

In an experiment discussed by [Kammers *et al.* \(2009b\)](#), the occluded right index finger and the visible index finger of a rubber hand were stroked either synchronously (illusion condition) or asynchronously (control condition). After this stimulation period, one of five perceptual localization responses was collected for $N = 14$ subjects. The following constrained models were of interest

- $\mathcal{M}_1 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$
- $\mathcal{M}_2 : \mu_1 = \mu_2 < \mu_3 = \mu_4 = \mu_5$
- $\mathcal{M}_3 : \mu_1 = \mu_2 < \mu_3 = \mu_4 < \mu_5,$

where μ_p is the measurement mean of the perceived location of the occluded hand, for $p = 1, \dots, 5$. Under model \mathcal{M}_1 , the location of the occluded hand is unaffected by the amount of new proprioceptive information. Under model \mathcal{M}_2 , the perceived location of the occluded hand is affected when moving the occluded hand, although it is independent of the amount of movements. Under model \mathcal{M}_3 , the perceived location of the hand is influenced by the number of movements of the occluded hand.

Repeated measurements 2: CONAMORE

The conflict and management of relationships study (CONAMORE; [Meeus *et al.* 2004](#)) is an ongoing longitudinal study in the Netherlands. The goal of the study is to assess relationships of adolescents with parents and best friend as well as the emotional states of the participants. In this example, the data consisted of $N = 340$ Dutch participants who were measured once

a year in a period of five years (from the age of 13 until 17). This was a part of a larger data set described in [Selfhout *et al.* \(2009\)](#).

The outcome variable was self-reported satisfaction with peer relations. There were two groups: boys and girls (coded by $j = 1$ and 2 in `data.txt`, respectively) and one time-varying covariate “generalized anxiety disorder” which was measured at each wave. Four models were analyzed,

- $\mathcal{M}_1 : \mu_{g13} = \mu_{b13}, \mu_{g14} = \mu_{b14}, \mu_{g15} = \mu_{b15}, \mu_{g16} = \mu_{b16}, \mu_{g17} = \mu_{b17}$
- $\mathcal{M}_2 : \mu_{g13} > \mu_{b13}, \mu_{g14} > \mu_{b14}, \mu_{g15} > \mu_{b15}, \mu_{g16} > \mu_{b16}, \mu_{g17} > \mu_{b17}$
- $\mathcal{M}_3 : 0 < \mu_{g13} - \mu_{b13} < \mu_{g14} - \mu_{b14} < \mu_{g15} - \mu_{b15} < \mu_{g16} - \mu_{b16} < \mu_{g17} - \mu_{b17}$
- $\mathcal{M}_4 : \mu_{g13} = \mu_{b13}, \mu_{g14} = \mu_{b14}, 0 < \mu_{g15} - \mu_{b15} < \mu_{g16} - \mu_{b16} < \mu_{g17} - \mu_{b17},$

where μ_{gp} and μ_{bp} is the mean of self-reported satisfaction of girls and boys at the age of p adjusted for generalized anxiety disorder, respectively. Model \mathcal{M}_1 states that there is no difference in self-reported satisfaction between boys and girls at each wave. Model \mathcal{M}_2 states that girls score their satisfaction with peer higher than boys at each wave. Model \mathcal{M}_3 states that girls score higher on self-reported satisfaction than boys and that the difference increases at each wave. Finally, model \mathcal{M}_4 states that that there is no difference between boys and girls at the age of 13 and 14, but from the age of 15 girls score higher than boys and the difference increases at each wave.

Multivariate one-sided testing: Cell counts of HIV-positive newborn infants

In this example, researchers were interested whether ritonavir therapy had a positive effect on the immunoreconstruction of 36 HIV-positive newborn infants ([Sleasman *et al.* 1999](#)). The analysis was limited to the cell counts of CD45RA T and CD45RO T, which were measured at birth and after 24 weeks of therapy ([Larocque & Labarre, 2004](#)). The models under investigation were

- $\mathcal{M}_1 : \boldsymbol{\mu} = \mathbf{0}$
- $\mathcal{M}_2 : \boldsymbol{\mu} > \mathbf{0}$

where $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top$ which contain the mean differences between the cell counts (after 24 weeks minus at birth) of the two cell types, and $\mathbf{0} = (0, 0)^\top$.

Multivariate analysis of variance: Does vinylidene fluoride cause liver damage?

Vinylidene fluoride is suspected of causing liver damage. To investigate this, 4 groups of 10 male Fischer-344 rats received different dosages of this substance by inhalation. Three serum enzyme levels (SDH, SGOT, and SGPT), which often indicate liver damage, were measured for each rat. The data, printed in ([Silvapulle and Sen 2005](#), p. 10), appeared in a report of Litton Bionetics Inc. (1984).

The data can be analyzed with a MANOVA model. When denoting $\boldsymbol{\mu}_j = (\mu_{j1}, \mu_{j2}, \mu_{j3})^\top$ with μ_{jp} the mean of the p th serum level of the j th group, for $j = 1, \dots, 4$, where the dosage of vinylidene fluoride increases from group 1 to 4, the following two models were of interest:

- $\mathcal{M}_1 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
- $\mathcal{M}_2 : \mu_1 < \mu_2 < \mu_3 < \mu_4$

Model \mathcal{M}_1 assumes that vinylidene fluoride has no influence on the three serum levels. Model \mathcal{M}_2 assumes that each serum level increases when increasing the dosage of vinylidene fluoride.

Multivariate regression: Cigarette burn and sugar percentage

The data that was provided by (Anderson and Bancroft 1952, p. 205) consisted of chemical components of 25 tobacco leaf samples consisting of six explanatory variables: nitrogen (x_1), chlorine (x_2), potassium (x_3), phosphorus (x_4), calcium (x_5), and magnesium (x_6), and two dependent variables: rate of cigarette burn in inches per 1,000 seconds (y_1) and per cent sugar in the leaf (y_2). After standardizing the dependent variables and covariates, the following models with inequality and equality constraints on the standardized regression coefficients α_{kp} with $k = 1, \dots, 6$ and $p = 1, 2$ (inspired by the results of Bedrick and Tsai 1994) were analyzed

- $\mathcal{M}_1 : \alpha_{11} = \dots = \alpha_{61}, \alpha_{12} = \dots = \alpha_{62}$
- $\mathcal{M}_2 : 0 < (\alpha_{11}, \alpha_{31}, -\alpha_{41}, \alpha_{51}, -\alpha_{61}) < -\alpha_{21}$, and
 $0 < (\alpha_{12}, \alpha_{32}, \alpha_{42}, \alpha_{52}, -\alpha_{62}) < \alpha_{22}$,
- $\mathcal{M}_3 : 0 = \alpha_{11} = \alpha_{41} = \alpha_{51} = \alpha_{61} < \alpha_{31} < -\alpha_{21}$,
 $0 = \alpha_{32} = \alpha_{42} = \alpha_{52} = \alpha_{62} < -\alpha_{12} < \alpha_{22}$.

Model \mathcal{M}_1 assumes all standardized regression coefficients corresponding to a dependent variable to be equal, model \mathcal{M}_2 assumes that chlorine dominates the other explanatory variables in predicting the two dependent variables and that all standardized regression coefficients are either smaller than zero or larger than zero, and model \mathcal{M}_3 assumes that each dependent variable is explained by only two specific explanatory variables where chlorine is dominant over the other explanatory variable.

6. Conclusions

The program that was discussed in this paper can be used to determine the Bayes factor of a multivariate normal linear model with equality and/or inequality constraints between the parameters against the unconstrained model. The outcome can be interpreted as the relative amount of evidence that is provided by the data in favor of the constrained model versus the unconstrained model. Due to the use of the CECPP (Mulder *et al.* 2009), which is based on certain linear restrictions and minimal training samples, the obtained Bayes factors appropriately balance between model fit and complexity for a broad class of equality and inequality constrained models.

A user-friendly interface (Appendix A) can be downloaded from the website <http://tinyurl.com/informativehypotheses>, which will allow researchers who are not familiar with the technical details of statistical modeling to use the proposed methodology. Also a stand-alone version (Appendix B) can be downloaded from this website. Because the Bayes factor of each

constrained model against the unconstrained model is computed separately in the stand-alone version, it is important that the same set of restrictions (Section 3.2, Appendix B), and therefore the same CECPP, is used for each computation. In the interface, this is automatically done. The interface also provides posterior model probabilities of all models under investigation. Based on the posterior model probabilities, the constrained models can be compared directly with each other concerning the amount of evidence provided by the data for each constrained model.

Acknowledgments

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

References

- Anderson RL, Bancroft TA (1952). *Statistical Theory in Research*. McGraw-Hill.
- Bartlett M (1957). “A Comment on D.V. Lindley’s Statistical Paradox.” *Biometrika*, **44**, 533–534.
- Bedrick EJ, Tsai CL (1994). “Model Selection for Multivariate Regression in Small Samples.” *Biometrics*, **50**, 226–231.
- Berger JO, Pericchi LR (1996). “The Intrinsic Bayes Factor for Model Selection and Prediction.” *Journal of American Statistical Association*, **91**, 109–122.
- Gelfand AE, Smith AFM, Lee T (1992). “Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling.” *Journal of the American Statistical Association*, **87**, 523–532.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis*. 2nd edition. Chapman & Hall, London.
- Hoijtink H, Klugkist I, Boelen PA (2008). *Bayesian Evaluation of Informative Hypotheses*. Springer-Verlag, New York.
- Huber P (1981). *Robust Statistics*. John Wiley & Sons.
- Jeffreys H (1961). *Theory of Probability*. 3rd edition. Oxford University Press, New York.
- Judge GG, Griffiths WE, Hill RC, Lee TC (1980). *The Theory and Practice of Econometrics*. John Wiley & Sons.
- Kammers MPM, de Vignemont F, Verhagen L, Dijkerman HC (2009b). “The Rubber Hand Illusion in Action.” *Neuropsychologica*, **47**, 204–211.
- Kammers MPM, Mulder J, de Vignemont F, Dijkerman HC (2009). “The Weight of Representing the Body: Addressing the Potentially Indefinite Number of Body Representations in Healthy Individuals.” *Experimental Brain Research*, **204**, 333–342.

- Kass RE, Raftery AE (1995). “Bayes Factors.” *Journal of American Statistical Association*, **90**, 773–795.
- Klugkist I, Hoijtink H (2007). “The Bayes Factor for Inequality and about Equality Constrained Models.” *Computational Statistics & Data Analysis*, **51**, 6367–6379.
- Laudy O (2006). *Bayesian Inequality Constrained Models for Categorical Data*. Ph.D. thesis, Utrecht University.
- Meeus W, Akse J, Branje S, Bogt TT, Engels R, et al CF (2004). “Codebook Conamore: Conflicts and Management of Relationships.” Utrecht University, Faculty of Social Sciences: Unpublished manuscript.
- Mulder J, Hoijtink H, Klugkist I (2010). “Equality and Inequality Constrained Multivariate Linear Models: Objective Model Selection Using Constrained Posterior Priors.” *Journal of Statistical Planning and Inference*, **140**, 887–906.
- Mulder J, Klugkist I, van de Schoot A, Meeus W, Selfhout M, Hoijtink H (2009). “Bayesian Model Selection of Informative Hypotheses for Repeated Measurements.” *Journal of Mathematical Psychology*, **53**, 530–546.
- Pérez JM, Berger JO (2002). “Expected Posterior Prior Distributions for Model Selection.” *Biometrika*, **89**, 491–511.
- Press SJ (2005). *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. 2nd edition. Krieger, Malabar.
- Selfhout M, Branje S, Meeus W (2009). “Development Trajectories of Perceived Friendship Intimacy, Constructive Problem Solving, and Depression from Early to Late Adolescence.” *Journal of Abnormal Child Behavior*, **37**, 251–264.
- Silvapulle MJ, Sen PK (2005). *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. 2nd edition. John Wiley & Sons.
- Sleasman JW, Nelson RP, Goodenow MM, Wilfert D, Hutson A, Basslaer M, Zuckerman J, Pizzo PA, Mueller BU (1999). “Immunoreconstitution After Ritonavir Therapy in Children with Human Immunodeficiency Virus Infection Involves Multiple Lymphocyte Lineages.” *Journal of Pediatrics*, **134**, 597–606.
- van Wesel F, Hoijtink H, Klugkist I (2011). “Choosing priors for inequality constrained normal linear models: Methods based on training samples.” *Scandinavian Journal of Statistics*, **38**(4), 666–690.
- Visual Numerics (2003). *IMSL Fortran Numerical Library*.
- Wetzels R, Grasman RPPP, Wagenmakers EJ (2010). “An Encompassing Prior Generalization of the Savage-Dickey Density Ratio Test.” *Computational Statistics & Data Analysis*, **54**(9), 2094–2102.

A. User manual of the BIEMS interface

The **BIEMS** interface can be obtained from <http://tinyurl.com/informativehypotheses>. After downloading and unzipping the folder containing this software package, the executable **BIEMS.exe** will be found after opening the folder. The purpose of this appendix is to describe how to use the interface. Only a data file **data.txt** needs to be provided, all the other information can be entered using the the **BIEMS** Windows interface. Execution of **BIEMS** will render output that can be viewed within the Windows interface and will be placed in the subfolder **BIEMS_calFBF**.

A.1. Data

The data in **data.txt** should be formatted as whitespace- or tab-separated values without header and with no missing values, where each column corresponds to a variable. **BIEMS** uses all variables in the data file, so only the variables of interest should be included. In addition, the variables in the data file need to be in the correct order. Consider an example with three dependent variables, one explanatory variable, one time-varying covariate, and two groups (i.e., using the notation introduced in Section 2, $P = 3$, $K = 1$, $U = 1$, $J = 2$). As was described in Sections 2.1 and 2.2, the observed variables in the data must ordered in **data.txt** as follows

y_{i1}	y_{i2}	y_{i3}	x_{i1}	w_{i11}	w_{i12}	w_{i13}	group
8.3	8.6	12.2	2.0	2.7	2.5	2.3	1
3.0	5.4	-1.0	-0.3	3.0	1.3	-0.5	2
7.0	9.1	9.6	1.7	2.7	2.2	1.4	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
3.3	-2.8	1.3	-1.3	3.8	-0.1	1.5	2

that is, first the dependent variables, then the explanatory variables, than the time-varying explanatory variables and finally the variable containing group membership. The group column is required, although it does not need to be sorted. If there is no grouping in the data, it should contain all 1's.

A.2. Step 1: Data selection

After clicking on **BIEMS.exe** the opening screen (Figure 5) will unfold. Clicking the button **Select** or **Generate** that can be found under the label **Step 1** in the upper left hand corner of the screen, renders the choice to generate data (which is discussed in the last section of this manual) or to import your own data. After importing your own data, these will be displayed in a new screen (Figure 6). In this screen you have to provide the **Number of dependent variables**, **Number of explanatory variables** and **Number of time varying explanatory variables**. After pressing **Ok**, a screen (Figure 7) will pop up that allows you to specify the constrained models of interest. This screen will be discussed in the next section.

A.3. Step 2: Specifying models

In the upper left hand corner of the model specification screen (Figure 7) you find a frame which is entitled **1: Compose**. The tool in this section allows you to specify model constraints.

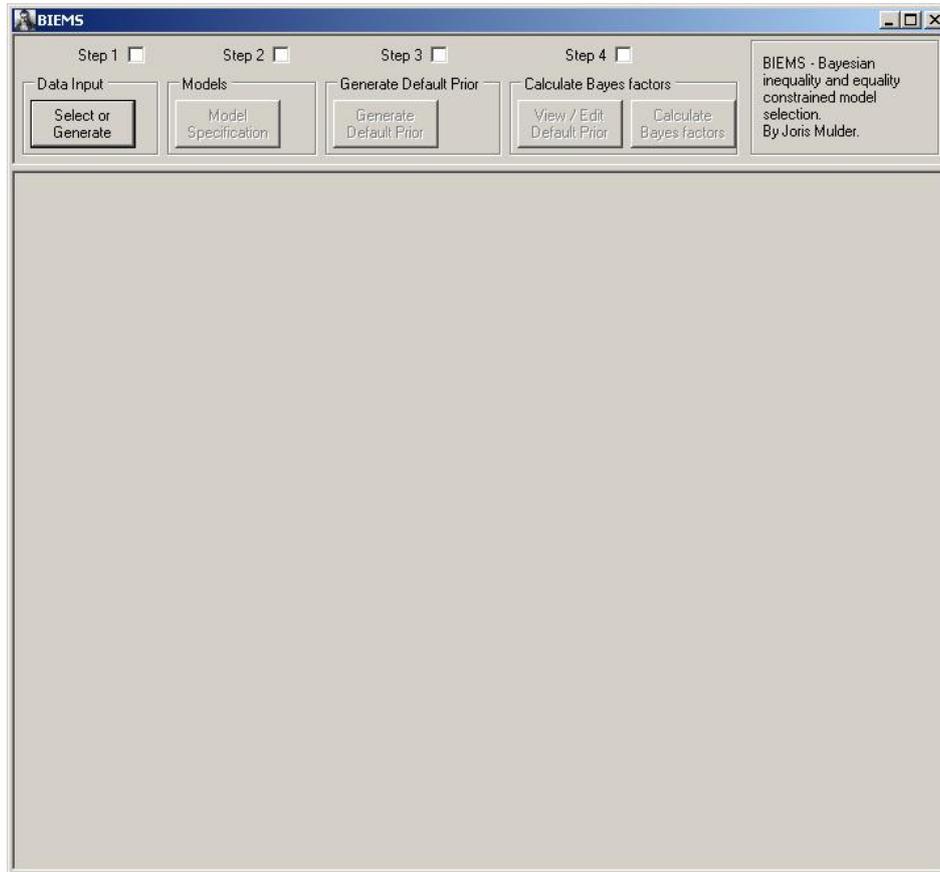


Figure 5: Opening screen.

If a model of interest is, for example, $\mu(1,1) > \mu(2,1) > \mu(3,1) > \mu(4,1)$ with respect to four means in a one-way ANOVA, this can be specified by subsequently entering:

- $\mu(1,1) > \mu(2,1)$ followed by a click on **Add to list**,
- $\mu(2,1) > \mu(3,1)$ followed by a click on **Add to list**,
- $\mu(3,1) > \mu(4,1)$ followed by a click on **Add to list**.

In the bottom left hand corner of the screen displayed in Figure 7, there is a frame 2: **List of Inequalities and Equalities**. Clicking on **Define as Model** places the composed model in the top right hand corner in the frame **List of Models**. Once this has been done, another model can be specified using the tool in the frame 1: **Compose**. To provide some examples of inequality and equality constraints one can specify:

- $\mu(1,1) - \mu(2,1) = \mu(3,1) - \mu(4,1)$,
- $\mu(1,1) - 2 * \mu(2,1) + \mu(3,1) > 0$,
- $\mu(1,1) - \mu(2,1) > 2.5$,
- $3 * \mu(1,1) - \mu(2,1) + 2 > 0$.

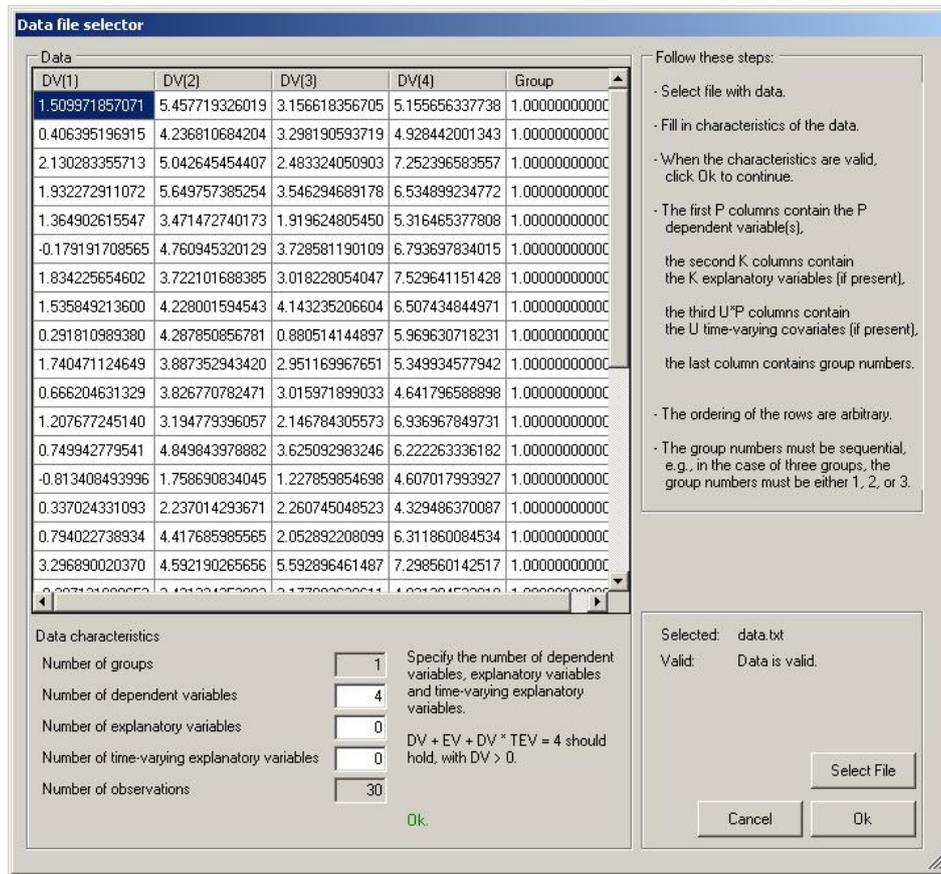


Figure 6: Data file selector screen. The model must be specified by setting the appropriate number of dependent variables, number of explanatory variables, and number of time-varying explanatory variables.

Note that the **Edit** and **Delete** buttons in the frames 2: **List of Inequalities and Equalities** and 3: **List of Models** can be used to modify or delete constraints in frame 2 and models in frame 3, respectively. As soon as the models of interest are specified the default prior distribution can be generated by clicking the button **Generate Default Prior** which can be found in the top of the screen. Before discussing this, some additional features in the model specification screen will be elaborated.

A.4. Additional features in the model specification screen

In the bottom right hand corner of the model specification screen (Figure 7), the frame 4: **Settings** can be found. Here the following default values for the computation of Bayes factors and posterior model probabilities can be modified:

- Ticking the circle in front of **Specify restrictions manually** allows manual specification of the restrictions used to construct the prior distribution using the tool in frame 1: **Compose** that was previously used to compose models. The restrictions that are used to specify the prior distribution are discussed in Section 3.2 of the paper.

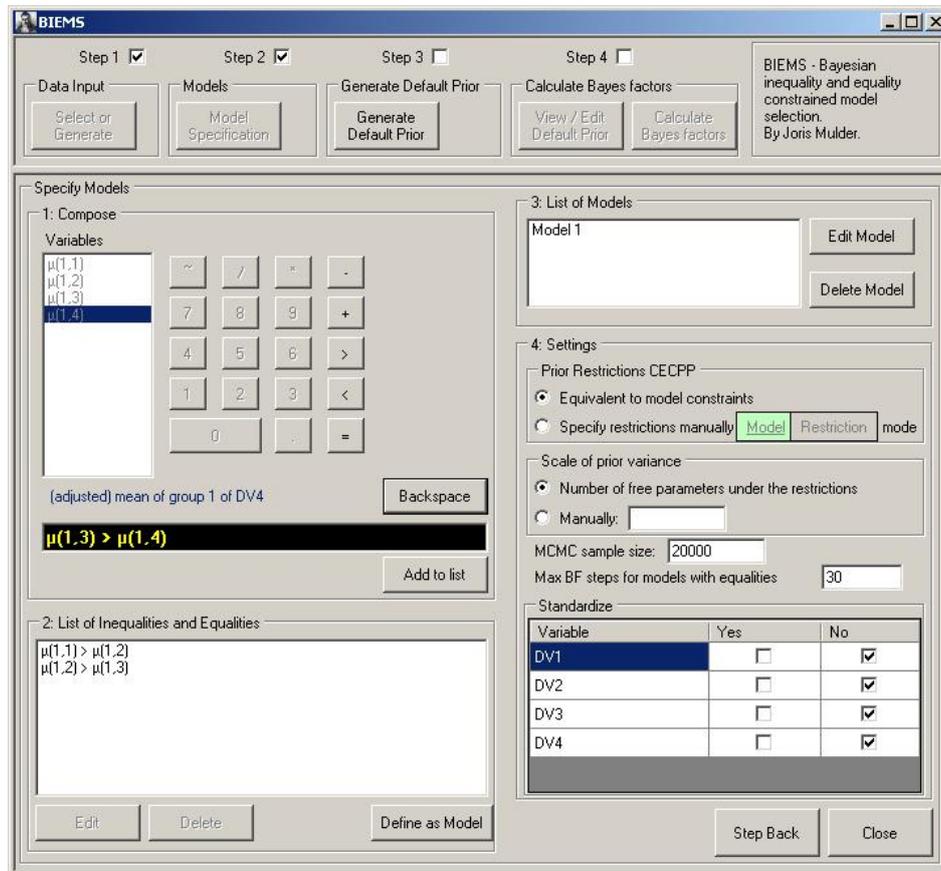


Figure 7: Screen where the equality and inequality constrained models can be specified.

- The scale of the prior variance can manually be adjusted by ticking the circle in front of **Manually**. The desired scale can be entered in the box located on the right of **Manually**. The default value of the scale of the prior distribution is discussed in Section 3.3.
- The default MCMC **sample size** can be adjusted. The MCMC sample size is discussed in Section 4.1 and in Section 4.2.
- The default number of **Max BF steps** can be adjusted. This number of steps is denoted by A in Section 4.3.

The last item, which can be found in the lower right hand corner of this screen, is important. Here it can be indicated whether or not the dependent variables and the (time-varying) explanatory variables should be standardized. Note that regression coefficients in a (multivariate) multiple regression are only comparable if both the dependent and independent variables are standardized (Section 5.2). Note also that in the context of an (M)ANCOVA, i.e., there are groups and continuous explanatory variables in the model, the continuous predictors have to be standardized (Section 5.4).

The button **Step Back** brings you back to the previous screen and deletes the information in the current screen. The button **Close** will end the current session with **BIEMS**.

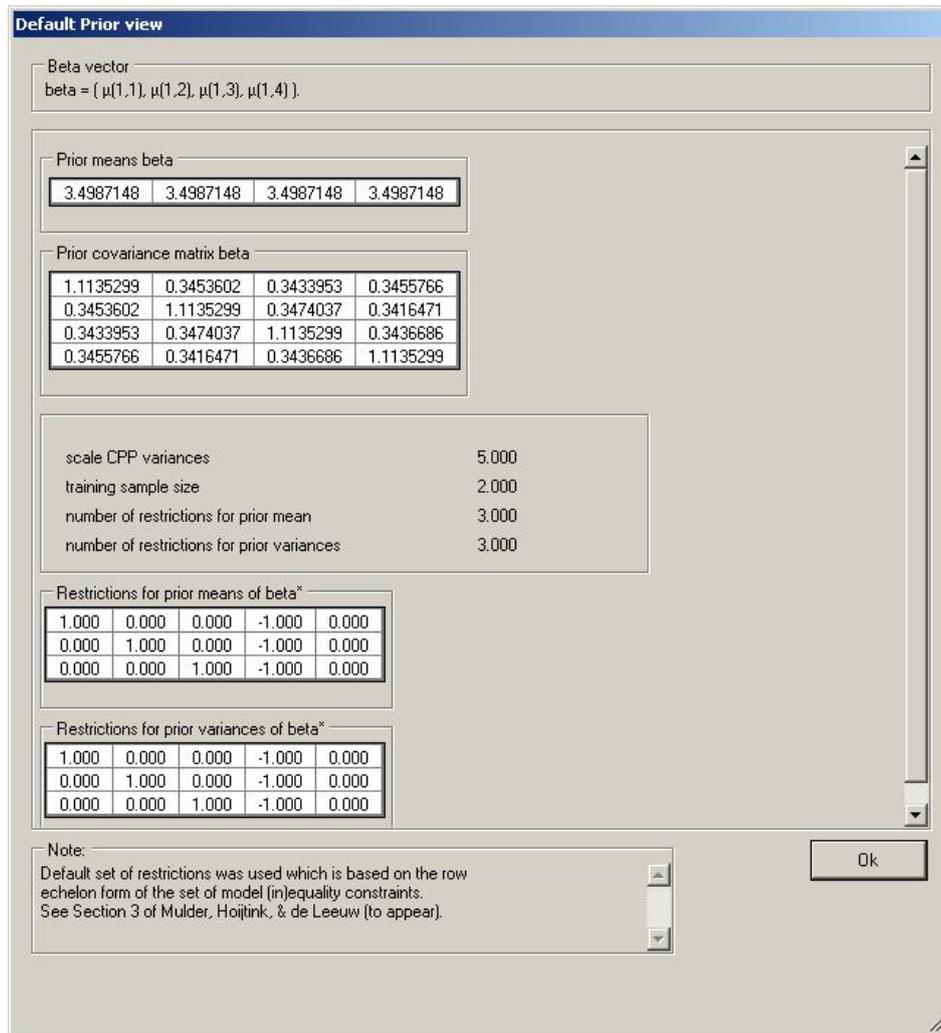


Figure 8: Default prior view which displays the properties and hyperparameters of the conjugate expected constrained posterior prior (CECPP).

A.5. Step 3: Generate the default prior distribution

As soon as the models of interest are specified, the default prior distribution can be generated by clicking the button **Generate Default Prior** which can be found in the top of the screen under **Step 3**. A screen will open (Figure 8) showing the prior means and covariance matrix, prior scale and training sample size, and matrices containing the restrictions used for the computation of the prior means and variances. More information with respect to the content of this screen can be found in Sections 3.3 and 3.4. The screen can be closed by clicking the **Ok** button. Note however that it is possible to change the prior means and prior covariance matrix by typing new values for the old values.

A.6. Step 4: View/edit default prior

Pressing the button **View/Edit Default Prior** which can be found in the top of the screen under **Step 4**, renders a return to the screen discussed in the previous section.

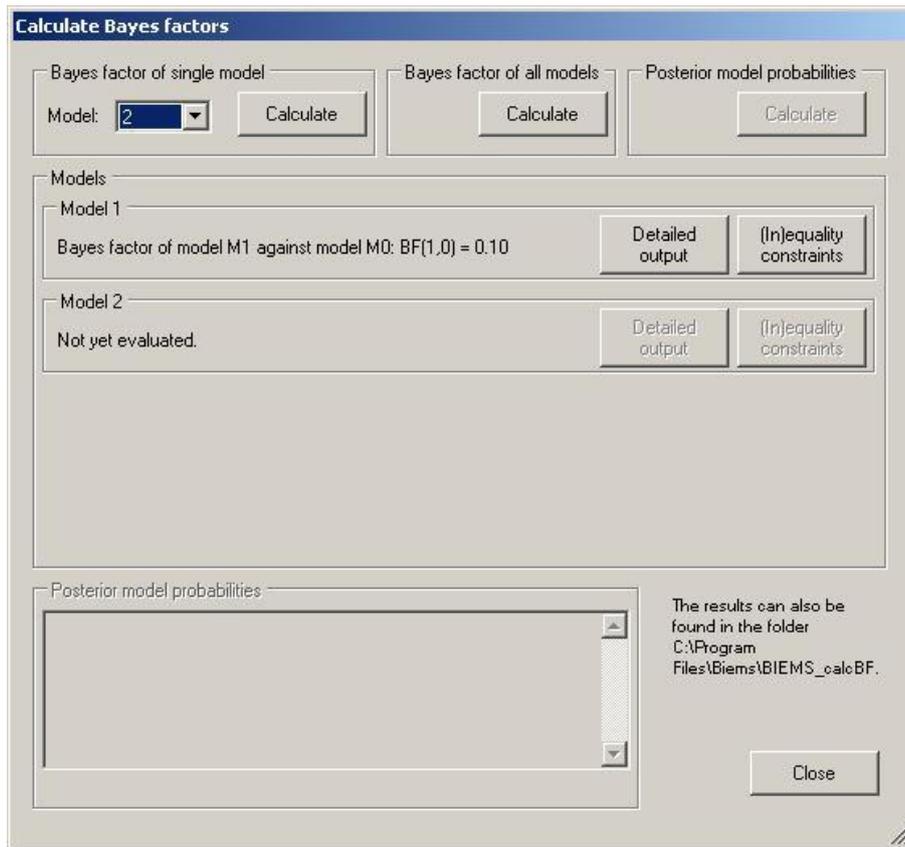


Figure 9: Screen where Bayes factors and posterior model probabilities are computed.

A.7. Step 4: Calculate Bayes factors

Pressing the button `Calculate Bayes factors`, which can be found in the top of the screen under `Step 4`, leads to a screen (Figure 9) where the Bayes factors are calculated. These will be displayed when the MCMC computation is done. For each model, detailed output can be inspected by clicking the buttons `Detailed output`. Furthermore, by pressing `(In)equality constraints`, the (in)equality constraints of each model can be viewed. The main features of the output will be discussed in the next section. Finally, the posterior model probabilities will be computed after pressing `Calculate` in the appropriate frame. These will be displayed in the frame below.

A.8. Output

The results of an analysis are given in the file `output_BIEMS.txt`. At the top of the file under the label `***Bayes factor***`, the Bayes factor for the constrained model \mathcal{M}_t against the unconstrained model \mathcal{M}_0 is given (Section 4), along with its standard error (Section 4.2):

Bayes factor of the constrained model M_t versus the unconstrained model M_0 :

`BF_{t,0} = 69.76646`

`standard error BF = 5.669907`

If equality constraints are used for the specification of \mathcal{M}_t , the Bayes factors for the sub steps (Section 4.3) are given as well. The values in the `subBFs` column are the Bayes factors for the constrained model in each step relative to the constrained model in the previous step. The `current BF` values are the estimated Bayes factors \hat{B}_{t0} at that particular step. In the example below, the `subBF` values can be seen to gradually approach the number 1.0 which implies that convergence is established.

subBFs	current BF
15.56523	15.56523
2.633227	40.98678
⋮	⋮
1.064050	69.54861
1.018774	70.85430
0.9846468	69.76646

If a model only contains inequality constraints, the estimated model complexity c_t and estimated model fit f_t (Section 4.1) are provided, as well as the Bayes factor for the model against its complement (Section 4). Under the label `***hits/BF in substeps***`, the details of the computation of f_t , c_t and the Bayes factor in each substep are provided (Section 4.1 under the label ‘More efficient estimation for small c_t ’).

In the output under the label `***Descriptive Statistics***`, the maximum likelihood estimate of the parameter matrix \mathbf{B} and the least squares estimate of $\mathbf{\Sigma}$ are given (Section 2.1). Under the label `***posterior***`, the posterior means and variances of \mathbf{B} are displayed. Under the label `***Prior***`, the CECPP mean β_0^{EC} is displayed in its matrix form \mathbf{B}_0^{EC} , followed by CECPP variances which are the diagonal elements of the prior covariance matrix $\mathbf{\Psi}_0^{EC}$ in (17), and finally the complete covariance matrix $\mathbf{\Psi}_0^{EC}$ is provided. These hyperparameters are ordered in essentially the same way as the matrix \mathbf{B} , e.g.,

```
CECPP mean of B
      1.294  1.294  1.294  (group 1)
      1.294  1.294  1.294  (group 2)
      4.038  4.692  4.557  (covariate 1)
      1.354  2.134  2.808  (time-varying covariate 1)
```

in the case of three dependent variables, two groups, one explanatory variable and one time varying covariate. Note that the row containing the estimates for the regression coefficients of the time-varying covariate correspond with the estimates of the diagonal elements of the matrix \mathbf{G}_1 in Section 2.2. Finally, under the label `***Prior Restrictions***`, the restrictions that were used for the computation of the hyperparameters of the constrained posterior priors (CPPs) are presented. First, the restrictions for the CPP means are presented, followed by the restrictions for the CPP variances. These restrictions are elaborated in Sections 3.2 and 3.3.

A.9. Step 1: Data generation

After clicking on `BIEMS.exe`, the opening screen (Figure 5) will unfold. Clicking the button `Select` or `Generate` renders the choice to generate data or to import your own data. If `Generate` is chosen, a screen will open (Figure 10) that allows the specification of a population from which data will be generated. The population is described by the multivariate normal linear model presented in Equation 1 in this paper, extended with a multivariate normal

Generate Data

Enter Settings

Number of dependent variables: 2

Number of groups: 4

Number of explanatory variables: 1

ISeed: -1

Edit Settings Use Settings

Group sizes | Error covariance matrix specification | Group means/regression coefficients | Explanatory variables

	DV1	DV2
Gr1	1	0
Gr2	2	2
Gr3	2	2
Gr4	4	2
EV1	5	9

Generate exact data

Generate

Generator Results

Cancel

Ok

Figure 10: Data generator screen.

distribution for the explanatory variables in each group.

In the opening screen, the Number of dependent variables, Number of Groups and Number of explanatory variables have to be given. By clicking the button Use Settings, a frame will be opened where the following characteristics must be given:

- The sample size of each group;
- The error covariance matrix Σ which is introduced a few lines below Equation 1;
- The group means M and the regression coefficients A which can be found in Equation 1 in this paper;
- The means and covariance matrix of the explanatory variables in each group.

After all parameters are specified, two options are available: simulate a data set from this population or generate a data set from the population that has sample parameters of the error covariance matrix, group means, regression coefficients, means and covariance matrix of the explanatory variables that are equal to the population parameters. If the latter is desired, the box labeled **Generate exact data** has to be ticked before clicking the button **Generate**. Data generated from a population with known properties can be useful to run tests with **BIEMS**, for example, to get an impression of the sample sizes needed in order to be able to evaluate the models of interest.

B. User manual of BIEMS (stand-alone version)

The purpose of this appendix is to describe the practical details of the stand-alone version of **BIEMS** which can be downloaded from the web site <http://tinyurl.com/informativehypotheses>. Unlike the interface, this version only computes the Bayes factor of one constrained model against the unconstrained model. It is not possible to specify multiple constrained models. For this reason, when more than one model is under investigation, say \mathcal{M}_1 and \mathcal{M}_2 , it is important that the same linear restrictions and therefore the same CECPP is used when computing the Bayes factors of each constrained model against the unconstrained model, B_{10} and B_{20} . Subsequently, the two constrained models can directly be compared with each other by calculating the Bayes factor of \mathcal{M}_1 against \mathcal{M}_2 by $B_{12} = B_{10}/B_{20}$.

A running example will be used throughout this appendix for illustration. The **BIEMS** program consists of one executable file `biems.exe`, a data file `data.txt`, a settings file `input_BIEMS.txt` and three further input files containing the constraints of the model and the restrictions on the prior necessary for calculating the Bayes factor of the constrained model against the unconstrained model (`equality_constraints.txt`, `inequality_constraints.txt` and `restriction_matrix.txt`). To calculate the Bayes factors of a model, set the model specification by modifying these text files with a plain text editor and run `biems.exe`. This will generate a file `output_BIEMS.txt` containing the output for the model. Note that this will overwrite previous versions of `output_BIEMS.txt` when present.

B.1. Data

The data in `data.txt` should be formatted as whitespace- or tab-separated values without header and with no missing values, where each column corresponds to a variable. **BIEMS** uses all variables in the data file, so only the variables of interest should be included. In addition, the variables in the data file need to be in the correct order. Throughout this appendix we consider an example with three dependent variables, one explanatory variable, one time-varying covariate, and two groups (i.e., $P = 3$, $K = 1$, $U = 1$, $J = 2$ when using the notation of Section 2.1 and 2.2). The observed variables in the data must ordered in `data.txt` as follows

y_{i1}	y_{i2}	y_{i3}	x_{i1}	w_{i11}	w_{i12}	w_{i13}	group
8.3	8.6	12.2	2.0	2.7	2.5	2.3	1
3.0	5.4	-1.0	-0.3	3.0	1.3	-0.5	2
7.0	9.1	9.6	1.7	2.7	2.2	1.4	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
3.3	-2.8	1.3	-1.3	3.8	-0.1	1.5	2

The group column is required, although it does not need to be sorted. If there is no grouping in the data, it should contain all 1's.

B.2. General settings

The general settings for **BIEMS** are specified in `input_BIEMS.txt`, which is divided into five input lines. On the first input line, the main properties of the data are defined:

```

Input 1:      #DV      #cov      #Tcov      N      iseed
              3        1        1        80      629535

```

These are the number of dependent variables, the number of explanatory variables, the number of time-varying covariates and the number of observations, P , K , U and N respectively (Sections 2.1 and 2.2). The `iseed` value is the seed of the random number generator, which can be set to any positive integer. It can also be set to `-1`. In that case a different random seed is used each time.

The second input line contains the number of constraints and restrictions that are placed on the model. We consider a model with 3 inequality constraints and 4 equality constraints (given in the next section), i.e.,

```

Input 2:      #inequalities  #equalities  #restrictions
              3              4              -1

```

These values should correspond to the number of equality constraints and inequality constraints of the model defined in the files `equality_constraints.txt` and `inequality_constraints.txt`, respectively. When the value for `#restrictions` is set to `-1`, the default set of restrictions on the prior is used which is equivalent to the row echelon form of the set of model constraints (see also Section 3.2). When the restrictions are manually specified in the file `restriction_matrix.txt` (discussed in the following section), for instance when multiple constrained models are under investigation and the default sets of restrictions are not identical across the models, the number of restrictions should then be stated in `#restrictions`.

The third input line allows the user to change some default values used by **BIEMS**:

```

Input 3:      sample size      maxBF steps      scale of CPP variance
              -1              -1              -1

```

Each of these options are set to `-1`, in which case the program's defaults are used. The first option specifies the number of draws used to estimate the (conditional) probabilities for each set of two inequality constraints, which defaults to 20,000 (see also Section 4.1 and 4.2). The second option gives the maximum number of steps used to approximate the Bayes factor for equality constraints (Section 4.3). By default, **BIEMS** uses as many steps as it takes for the Bayes factors between adjacent steps to become approximately one. The third option gives the scale of the variances of the constrained posterior prior for β . By default this is equal to the number of free parameters under the restricted model (Section 3.3).

The fourth and fifth input lines allow the user to standardize the dependent and explanatory variables:

```

Input 4:      standardize DV/IV
              0        2

```

```

Input 5:      covariates to standardize (if standardize IV = 2)
              1 0

```

The first value on input line 4 indicates whether to standardize the dependent variables, the second value whether to standardize the explanatory variables and time-varying covariates

(0 = no, 1 = yes). The second value can also be set to 2, to indicate that only some of the explanatory variables and/or time-varying covariates should be standardized. This is specified on input line 5, which should then contain a line of 0's and 1's of length $K + U$, indicating for each explanatory variable and time-varying covariate whether it should be standardized or not. As specified above for the example, only the explanatory variable is standardized. When testing constraints on the regression coefficients, it is recommendable to standardize the dependent variable(s) and the (time-varying) explanatory variable(s), so that actually the constraints on the standardized regression coefficients are evaluated. Also in the case of a (M)ANCOVA, we recommend to standardize the explanatory variables.

B.3. Constraints and restrictions

The constraints of the model and restrictions on the prior are set in the files `equality_constraints.txt`, `inequality_constraints.txt` and `restriction_matrix.txt`. If `#restrictions` is set to `-1` in `input_BIEMS.txt`, the latter file does not need to be specified. The first two of these files contain the matrices $[\mathbf{Q}_t|\mathbf{q}_t]$ and $[\mathbf{R}_t|\mathbf{r}_t]$, as defined in Section 2.3. All three files have the same structure, with each row corresponding to a constraint or restriction.

For the example we have

$$\mathbf{B} = \begin{array}{ccc} & DV1 & DV2 & DV3 \\ \left(\begin{array}{ccc} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_{22} & \mu_{23} \\ \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \gamma_{11} & 0 & 0 \\ 0 & \gamma_{12} & 0 \\ 0 & 0 & \gamma_{13} \end{array} \right) & \left. \begin{array}{l} \text{group 1} \\ \text{group 2} \\ \text{expl. var. 1} \\ \text{time-var. cov. 1.} \end{array} \right\} \end{array}$$

The vectorization of this matrix along its columns results in the following parameter vector

$$\boldsymbol{\beta}^\top = (\mu_{11}, \mu_{21}, \alpha_{11}, \gamma_{11}, \mu_{12}, \mu_{22}, \alpha_{12}, \gamma_{12}, \mu_{13}, \mu_{23}, \alpha_{13}, \gamma_{13}).$$

The model under evaluation has the following constraints:

$$\begin{array}{ccccc} \mu_{11} & = & \mu_{12} & = & \mu_{13} \\ \wedge & & \wedge & & \wedge \\ \mu_{21} & = & \mu_{22} & = & \mu_{23}. \end{array}$$

Thus, it is expected that the group means do not change over time and the mean of group 2 is higher than the mean of group 1 at each time point. This is specified in the input files as follows (parameter names have been added here for convenience):

equality_constraints.txt												
μ_{11}	μ_{21}	α_{11}	γ_{11}	μ_{12}	μ_{22}	α_{12}	γ_{12}	μ_{13}	μ_{23}	α_{13}	γ_{13}	
1	0	0	0	-1	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	-1	0	0	0	0
0	1	0	0	0	-1	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	-1	0	0	0

and

inequality_constraints.txt												
μ_{11}	μ_{21}	α_{11}	γ_{11}	μ_{12}	μ_{22}	α_{12}	γ_{12}	μ_{13}	μ_{23}	α_{13}	γ_{13}	
-1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	-1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	-1	1	0	0	0

In this example, the default set of restrictions (i.e., `#restrictions = -1` in `input_BIEMS.txt`) can be used for constructing the conjugate expected-constrained posterior prior (CECPP). This is equal to the row echelon form of the combination of inequality and equality constraints of the model under evaluation. When the constraints in `inequality_constraints.txt` and `equality_constraints.txt` are combined and row operations are performed, the system is containing two rows with zeros. After removing these two rows, the following set of restrictions is used

restriction_matrix.txt												
μ_{11}	μ_{21}	α_{11}	γ_{11}	μ_{12}	μ_{22}	α_{12}	γ_{12}	μ_{13}	μ_{23}	α_{13}	γ_{13}	
1	0	0	0	0	0	0	0	0	-1	0	0	0
0	1	0	0	0	0	0	0	0	-1	0	0	0
0	0	0	0	1	0	0	0	0	-1	0	0	0
0	0	0	0	0	1	0	0	0	-1	0	0	0
0	0	0	0	0	0	0	0	1	-1	0	0	0,

which results in equal prior distributions for $\mu_{11}, \dots, \mu_{23}$ (Section 3).

B.4. Output

The results of an analysis are given in the file `output_BIEMS.txt`. At the top of the file the Bayes factor for the constrained model against the unconstrained model is given, along with its standard error (see Section 4.2):

Bayes factor of the constrained model M_t versus the unconstrained model M_0:

BF_{t,0} = 69.76646

standard error BF = 5.669907

If equality constraints were used, the Bayes factors for the substeps (see Section 3.3) are given as well. The values in the `subBF` column are the Bayes factors for the model in that step, relative to the model in the previous step. The `current BF` values are the estimated Bayes factors \hat{B}_{t0} for that step. For the example, the `subBF` values can be seen to gradually approach the final Bayes factor of 69.8.

subBFs	current BF
15.56523	15.56523
2.633227	40.98678
⋮	⋮
1.064050	69.54861
1.018774	70.85430
0.9846468	69.76646

If only inequality constraints were used, the estimated model complexity c_t and estimated model fit f_t are provided, as well as the Bayes factor for the model against its complement (Section 4). Following this are some more details for each of the substeps.

At the bottom of the file some descriptive statistics are given: the maximum likelihood (ML) estimate of the parameter matrix \mathbf{B} (when time-varying covariates are present the restricted ML estimate of \mathbf{B} is provided given that $\gamma^* = \mathbf{0}$ (Section 2.2)), the posterior means and variances of \mathbf{B} , as well as the means, variances and covariances for the conjugate expected-constrained posterior prior. The values given under *****prior***** correspond to the symbols in equation (5.15) as follows: under **CECPP means of B** are the values of β_0^{EC} (in matrix form \mathbf{B}), **CECPP variance of B** contains the CECPP variances of the elements in \mathbf{B} , and **CECPP covariance matrix of beta** represents the prior covariance matrix Ψ_0^{EC} of β . Parameters for these and the earlier descriptive statistics are ordered in essentially the same way as the matrix \mathbf{B} , e.g.,

```
CECPP mean of B
    1.294  1.294  1.294  (group 1)
    1.294  1.294  1.294  (group 2)
    4.038  4.692  4.557  (covariate 1)
    1.354  2.134  2.808  (time-varying covariate 1)
```

where the columns represent the measurements at three different time points. As a result of the CECPP restrictions, the prior means of the adjusted measurement means μ are identical in the above matrix. The complete matrix Ψ_0^{EC} is given as **CECPP covariance matrix of beta**, and the order of the elements in \mathbf{B} is identical to the order of the elements in $\beta = \text{vec}(\mathbf{B})$. Finally under *****Prior Restrictions*****, the restriction matrices are given that were used to obtain the CECPP means and the CECPP variances of β .

Affiliation:

Joris Mulder

Department of Methodology and Statistics

Tilburg University

5037 AB Tilburg, The Netherlands

E-mail: j.mulder3@uvt.nl

URL: <http://www.tilburguniversity.edu/webwijs/show/?anr=629387>