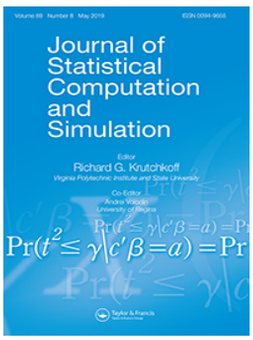

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Bain: a program for Bayesian testing of order constrained hypotheses in structural equation models

Xin Gu, Herbert Hoijtink, Joris Mulder & Yves Rosseel

To cite this article: Xin Gu, Herbert Hoijtink, Joris Mulder & Yves Rosseel (2019) Bain: a program for Bayesian testing of order constrained hypotheses in structural equation models, Journal of Statistical Computation and Simulation, 89:8, 1526-1553, DOI: 10.1080/00949655.2019.1590574

To link to this article: https://doi.org/10.1080/00949655.2019.1590574

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
2019, VOL. 89, NO. 8, 1526–1553
https://doi.org/10.1080/00949655.2019.1590574

Taylor & Francis
Taylor & Francis Group

Bain: a program for Bayesian testing of order constrained hypotheses in structural equation models

Xin Gu[a], Herbert Hoijtink[b,c], Joris Mulder[d] and Yves Rosseel[e]

[a]Department of Educational Psychology, East China Normal University, Shanghai, China; [b]Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands; [c]CITO Institute for Educational Measurement, Arnhem, The Netherlands; [d]Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands; [e]Department of Data Analysis, Ghent University, Ghent, Belgium

ABSTRACT

This paper presents a new statistical method and accompanying software for the evaluation of order constrained hypotheses in structural equation models (SEM). The method is based on a large sample approximation of the Bayes factor using a prior with a data-based correlational structure. An efficient algorithm is written into an R package to ensure fast computation. The package, referred to as Bain, is easy to use for applied researchers. Two classical examples from the SEM literature are used to illustrate the methodology and software.

ARTICLE HISTORY
Received 10 August 2017
Accepted 27 February 2019

KEYWORDS
Approximate Bayesian procedure; Bayes factors; order constrained hypothesis; structural equation model

1. Introduction

Applied researchers have become increasingly interested in the evaluation of order constrained hypotheses because the traditional null hypothesis is often not a realistic representation of the population of interest [[1],[2, pp. 79–81]]. In structural equation models, researchers may have explicit theories or expectations, for example, about the ordering of the relative effects of independent variables on a dependent variable or researchers may expect which indicator for a latent variable is dominant over the other indicators. These expectations can be represented by order constrained hypotheses among the model parameters. Order constrained hypotheses can be evaluated using either the frequentist approach by means of p values (see, e.g. [3,4]) or the Bayesian approach by means of Bayes factors (see, e.g. [5–7]). In this paper, the Bayes factor [8] is used as a criterion for assessing the hypotheses because p values can only reject a null hypothesis. Bayes factors on the other hand are able to measure the relative evidence in the data between multiple non-nested hypotheses containing order constraints [9]. For this reason, Bayes factors can be viewed as a more generally applicable tool for statistical hypothesis testing than classical p values.

During the past decade, Bayesian evaluation of hypotheses with order (or inequality) constraints on the parameters of interest has been studied for various statistical

CONTACT Xin Gu guxin57@hotmail.com Department of Educational Psychology, East China Normal University, Shanghai, China

models. Besides statistical theory development, these studies rendered software packages that can be used by applied researchers, see [5] for an overview. As a pioneer, Klugkist et al. [6] presented a Bayesian approach to evaluate analysis of (co)variance models (ANOVA or ANCOVA) with order constraints on the means. The study for ANOVA models was further developed by Kuiper and Hoijtink [10] for the comparison of means using both Bayesian and non-Bayesian methods. This research resulted in a software package `ConfirmatoryANOVA` [11]. Thereafter, Mulder et al. [12] extended the previous study to multivariate linear models (MANOVA, repeated measures, multivariate regression), which is implemented in the software package `BIEMS` [13]. Finally, Gu et al. [14] explored a general Bayesian procedure using a diffuse normal prior distribution with a diagonal covariance structure. Although this methodology provided reasonable default outcomes of the Bayes factor, the diagonal prior covariance structure can be criticized because the resulting Bayes factor is not invariant for linear one-to-one transformations of the data. The invariance property is important because it ensures that the relative evidence between two hypotheses, as quantified by the Bayes factor, does not depend on the arbitrary parameterizations of the model [15].

To illustrate the issue, consider three repeated measurements coming from a multivariate normal distribution, i.e. $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})^{\mathrm{T}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}}$ is a vector containing the measurement means and $\boldsymbol{\Sigma}$ is the measurement covariance matrix. Now assume we are interested in testing a monotonic increase of the means against an unrestricted alternative: $H_1 : \theta_1 < \theta_2 < \theta_3$ versus $H_u : \boldsymbol{\theta} \in \mathbb{R}^3$ where $\mathbb{R}^3$ denotes the 3-dimensional real vector space. A standard choice for the prior under H_1 is to use a truncation of the unconstrained prior under H_u in the order constrained space under H_1 (e.g. [6]). This results in the following expression of the Bayes factor:

$$B_{1u} = \frac{\Pr(\theta_1 < \theta_2 < \theta_3 \mid \mathbf{X}, H_u)}{\Pr(\theta_1 < \theta_2 < \theta_3 \mid H_u)}, \tag{1}$$

which corresponds to the ratio of the posterior probability that the constraints of H_1 are satisfied under H_u and the prior probability that the constraints of H_1 are satisfied under H_u. Now consider a very vague prior with a multivariate normal distribution for the measurement means under H_u with a diagonal covariance structure, $\pi_u(\boldsymbol{\theta}) = N(\mathbf{0}, \omega \mathbf{I}_3)$, where $\mathbf{I}_3$ is a 3-dimensional identity matrix and ω is chosen large enough so that the posterior probability in the numerator in (1) is virtually independent of the prior, say, $\omega = 10^6$. This prior results in a prior probability that the constraints hold that is equal to $\frac{1}{6}$ (for any choice of ω, see [6]), and thus, the Bayes factor is equal to the posterior probability that the measurement means increase multiplied by 6, i.e. $B_{1u} = 6 \times \Pr(\theta_1 < \theta_2 < \theta_3 \mid \mathbf{X}, H_u)$.

Now we consider a one-to-one transformation of the data where the first element corresponds to the difference between the first and second repeated measurement, the second element corresponds to the difference between the second and third repeated measurement, and the third element corresponds to the third repeated measurement, i.e. $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}) = (x_{i1} - x_{i2}, x_{i2} - x_{i3}, x_{i3})$. Again the transformed observations follow a multivariate normal distribution, say, $\mathbf{y}_i \sim N(\boldsymbol{\eta}, \boldsymbol{\Psi})$, where the first, second, and third element of $\boldsymbol{\eta}$ are equal to the first mean difference, the second mean difference, and the mean of the third observation. The equivalent hypothesis test in this parameterization comes down to $H_1 : \eta_1 < 0, \eta_2 < 0, \eta_3 \in \mathbb{R}^1$ versus $H_u : \boldsymbol{\eta} \in \mathbb{R}^3$. Similarly as in (1), the Bayes factor is

now given by

$$B_{1u} = \frac{\Pr(\eta_1 < 0, \eta_2 < 0 \mid \mathbf{Y}, H_u)}{\Pr(\eta_1 < 0, \eta_2 < 0 \mid H_u)}.$$

Again we consider independent normal priors for the mean parameters, i.e. $\pi_u(\boldsymbol{\eta}) = N(\mathbf{0}, \omega\mathbf{I}_3)$, with ω very large. In this case the prior probability that the constraints of H_1 hold under H_u equals $\frac{1}{4}$, and consequently, the Bayes factor is equal to the posterior probability of negative mean differences (which is equivalent to an increase of the measurement means) multiplied by 4, i.e. $B_{1u} = 4 \times \Pr(\eta_1 < 0, \eta_2 < 0 \mid \mathbf{Y}, H_u)$. Thus, the Bayes factor differs with a factor of $\frac{4}{6}$ for these two parameterizations, which is quite large. For larger dimensions with, say, 10 measurements, the violation will be even larger [15,16]. This is highly undesirable. To resolve this we present a new default prior resulting in a new Bayesian testing procedure for testing order constrained hypotheses in SEM which avoids this issue. The general idea is to let the prior covariance structure of the parameters of interest to depend on the covariance structure in the sample.

The second main contribution is the development of an efficient algorithm for computing the prior and posterior probability that a set of order (inequality) constraints hold, which are key quantities when computing Bayes factors. This contribution is needed because computing these probabilities as the proportion of draws satisfying the constraints can be very inefficient when the hypotheses contain many order constraints on the parameters of interest. In this case the posterior and prior probability that the constraints hold can be very small and therefore billions of draws may be needed in order to get accurate estimates of the probabilities and the resulting Bayes factors [5]. For this reason an efficient algorithm is presented that consists of roughly two steps. First, the probability of a set of order constraints is written as product of conditional probabilities. Second, the conditional probabilities are computed as the arithmetic mean of conditional probabilities which have analytic expressions in the Gibbs sampler. As will be seen this new algorithm is much more efficient than the use of the proportion of draws satisfying the constraints.

The algorithm is implemented into an R package referred to as Bain to ensure fast computation. For the computation of the Bayes factor using Bain the user only needs to provide the estimates of the parameters of interest and the inverse of the Fisher information matrix of the parameters which serves as the posterior covariance matrix. These statistics can for instance be obtained using the `lavaan` package [17] in `R` for the analysis of structural equation models (SEM). Other software, such as `Mplus`, can also be used to obtain these statistics, but here we will use `lavaan` as the basis for our analyzes because it is free.

In what follows, Section 2 shortly introduces SEM models and defines order constrained hypotheses. For the evaluation of order constrained hypotheses, the Bayes factor as a criterion is briefly introduced in Section 3. Subsequently, Section 4 specifies prior and posterior distributions which are the determinants of the Bayes factor. Thereafter, the procedure for the computation of Bayes factors is presented in Section 5 in which seven sub-sections describe the principles and algorithms used. To illustrate how to evaluate order constrained hypotheses using our programme, Section 6 analyzes two classic SEM models: confirmatory factor analysis and multiple regression models with latent variables. Finally, a user manual is provided in Appendix 2 such that researchers can use the implementation in Bain successfully for the analysis of their own data.

2. Order constrained structural equation models

2.1. Structural equation models

The structural equation model (SEM) mainly consists of two components, i.e. the measurement model which expresses the relations between latent variables and their indicators, and the structural model which expresses the relations between endogenous and exogenous (latent) variables, see for example [18]. The measurement model can be written by

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\epsilon}_y \\ \boldsymbol{x} &= \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\epsilon}_x \end{aligned} \tag{2}$$

where $\boldsymbol{y}$ and $\boldsymbol{x}$ denote the vectors of endogenous and exogenous observed variables, respectively, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ denote the vectors of endogenous and exogenous latent variables, respectively, $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are the corresponding matrices of factor loadings, and the measurement errors $\boldsymbol{\epsilon}_y$ and $\boldsymbol{\epsilon}_x$ have zero means and covariance matrices $\boldsymbol{\Psi}_{\epsilon_y}$ and $\boldsymbol{\Psi}_{\epsilon_x}$, respectively.

The structural model represents the relations among latent variables:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3}$$

where $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are matrices of regression coefficients, and $\boldsymbol{\delta}$ with mean of $\mathbf{0}$ and covariance matrix of $\boldsymbol{\Psi}_\delta$ is the error term. In addition,

$$\boldsymbol{\Phi}_\eta = (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}_\xi\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Psi}_\delta)(\mathbf{I}^{\mathrm{T}} - \mathbf{B}^{\mathrm{T}})^{-1}, \tag{4}$$

where $\boldsymbol{\Phi}_\eta$ and $\boldsymbol{\Phi}_\xi$ are the covariance matrices of the latent variables $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, respectively. Note that both $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ may contain observed variables if one wants to model the relationship between observed variables. This can be done by creating single-indicator latent variables (with a fixed factor loading of 1, and zero measurement error) corresponding to each observed variable.

The general framework of SEM is described by Equations (2) and (3) which can be specified using `lavaan` syntax [17] in `R`. As can be seen from (2), (3) and (4), the non-fixed elements in $\{\boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x, \mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Psi}_{\epsilon_y}, \boldsymbol{\Psi}_{\epsilon_x}, \boldsymbol{\Psi}_\delta, \boldsymbol{\Phi}_\xi\}$ of a specific SEM model can be collected in a parameter vector $\boldsymbol{\lambda}$. The density of the data is given by $f(\boldsymbol{X} \mid \boldsymbol{\lambda})$, where $\boldsymbol{X}$ denotes the data [19]. The distribution of data $\boldsymbol{X}$ is most often multivariate normal, though it could also involve multinomial distribution, et al. Furthermore, the non-fixed parameters can be divided into $\boldsymbol{\lambda} = \{\boldsymbol{\theta}, \boldsymbol{\zeta}\}$, where $\boldsymbol{\theta}$ denotes the target parameters that will appear in the order constrained hypotheses elaborated in the next section, and $\boldsymbol{\zeta}$ denotes the nuisance parameters that will not.

2.2. Order constrained hypotheses

Order constrained hypotheses express the expectations of researchers among the (standardized) target parameters in SEM. For example, hypothesis $H_1 : \theta_1 > \theta_2$ where θ_1 and θ_2 are the coefficients of the predictors ξ_1 and ξ_2, respectively, implies that the predictor ξ_1 is stronger than ξ_2. The general form of an order constrained hypothesis H_i is given by

$$H_i : \boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i, \tag{5}$$

where $\boldsymbol{R}_i$ is the restriction matrix containing order constraints, and $\boldsymbol{\theta}$ and $\boldsymbol{r}_i$ denote the target parameter vector and constant vector in H_i, respectively. We assume that the number

of constraints is K and the number of target parameters is J. Therefore, $\boldsymbol{R}_i$ is a $K \times J$ matrix, and the lengths of $\boldsymbol{\theta}$ and $\boldsymbol{r}_i$ are J and K, respectively. For instance, $H_2 : \theta_1 > \theta_2 > \theta_3$ is an example with $J = 3$ and $K = 2$, which leads to $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}}$ and an augmented matrix:

$$[\boldsymbol{R}_2 \mid \boldsymbol{r}_2] = \left[\begin{array}{ccc:c} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{array}\right].$$

The augmented matrix $[\boldsymbol{R}_i \mid \boldsymbol{r}_i]$ should be implemented as input of Bain.

The hypothesis H_i is often compared to an unconstrained hypothesis

$$H_u : \boldsymbol{\theta} \in \mathbb{R}^J, \tag{6}$$

where $\mathbb{R}^J$ denotes the J-dimensional real vector space, or to its complement

$$H_{i_c} : \text{not } H_i. \tag{7}$$

Furthermore, we can evaluate H_i against a competing hypothesis

$$H_{i'} : \boldsymbol{R}_{i'}\boldsymbol{\theta} > \boldsymbol{r}_{i'}. \tag{8}$$

The evaluation of these hypotheses can be conducted using Bayes factors, which will be elaborated in the next section.

When specifying order constrained hypotheses in SEM models, the target parameters may need to be standardized. For example, if hypothesis $H_1 : \theta_1 > \theta_2$ compares two regression coefficients to determine which predictor is stronger, then the coefficients θ_1 and θ_2 should be standardized to be comparable. The standardization of target parameters can be achieved by standardizing the observed and latent variables in SEM models. However, this manner might be criticized because the data is used twice, once for standardization and once for evaluation of the hypothesis [14]. The `lavaan` package [17] provides an alternative approach that can directly obtain estimates and covariance matrix of standardized target parameters. This paper uses the alternative standardization approach in lavaan. To keep the notation simple, in this paper $\boldsymbol{\theta}$ will be used to denote both unstandardized and standardized target parameters.

3. Bayes factor

The Bayes factor of H_i against H_u is defined as the ratio of two marginal likelihoods [5,8,20]:

$$BF_{iu} = \frac{m_i(\boldsymbol{X})}{m_u(\boldsymbol{X})} = \frac{\int\int f(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{\zeta})\pi_i(\boldsymbol{\theta}, \boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\zeta}}{\int\int f(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{\zeta})\pi_u(\boldsymbol{\theta}, \boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\zeta}}, \tag{9}$$

where $\pi_i(\boldsymbol{\theta}, \boldsymbol{\zeta})$ and $\pi_u(\boldsymbol{\theta}, \boldsymbol{\zeta})$ denote the prior distribution under H_i and H_u (will be specified in the next section), respectively, and $f(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{\zeta})$ denotes the density of $\boldsymbol{X}$ given $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ (see [19]). Furthermore, from Equation (9) it follows that the Bayes factor of H_i against H_{i_c} can be obtained as $BF_{ii_c} = BF_{iu}/BF_{i_cu}$, and the Bayes factor of H_i against $H_{i'}$ is $BF_{ii'} = BF_{iu}/BF_{i'u}$.

The Bayes factor BF_{iu} quantifies the relative evidence in the data in favour of hypothesis H_i against H_u. For example $BF_{iu} = 2$ indicates that the support in the data for H_i is

twice as large as the support for H_u. A general guideline for the interpretation of the Bayes factor is that $BF_{iu} \in (1, 3]$ indicates evidence for H_i that is not worth mentioning, and $BF_{iu} \in (3, 20]$, $BF_{iu} \in (20, 150]$ and $BF_{iu} > 150$ indicate positive, strong and very strong evidence for H_i, respectively [8]. Note that if $BF_{iu} < 1$ which implies evidence against H_i, the strength of this evidence is quantified using the rule above for the reciprocal of BF_{iu}. Furthermore, Bayes factors BF_{ii_c} and $BF_{ii'}$ can also be interpreted using the same rule. Although this rule renders a proposal to interpret the Bayes factor, it is not suggested using it strictly because this interpretation is a rough descriptive statement with respect to the standards of evidence, which could very well be modified based on the research context. For this reason users can judge by themselves when the evidence in the data is positive, strong or decisive in favour or against a hypothesis based on the observed Bayes factor.

Formula (9) can be simplified to [21]:

$$BF_{iu} = \frac{f_i}{c_i}, \tag{10}$$

where

$$c_i = \iint_{\theta \in \Theta_i} \pi_u(\boldsymbol{\theta}, \boldsymbol{\zeta})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\zeta} = \int_{\theta \in \Theta_i} \pi_u(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}, \tag{11}$$

called relative complexity [16], is the proportion of the prior distribution (specified in the next section) in agreement with H_i relative to H_u, and

$$f_i = \iint_{\theta \in \Theta_i} \pi_u(\boldsymbol{\theta}, \boldsymbol{\zeta} \mid \boldsymbol{X})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\zeta} = \int_{\theta \in \Theta_i} \pi_u(\boldsymbol{\theta} \mid \boldsymbol{X})\, \mathrm{d}\boldsymbol{\theta}, \tag{12}$$

called relative fit, is the proportion of the posterior distribution (specified in the next section) in agreement with H_i relative to H_u. Here $\Theta_i = \{\boldsymbol{\theta} \mid \boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i\}$ denotes the parameter space constrained by H_i, and $\boldsymbol{\zeta}$ is not constrained. The complexity implies how specific a hypothesis is, and the fit implies how much the data supports a hypothesis relative to H_u. The more specific the hypothesis, the less the complexity, while the more the support from the data, the larger the fit. The derivation of Equation (10) can be found in [16]. Equation (10) shows that the Bayes factor of an order constrained hypothesis H_i against an unconstrained hypothesis H_u can be represented as the ratio of the fit and complexity of H_i. This representation facilitates our development of the software for the evaluation of order constrained hypotheses.

Based on BF_{iu}, the Bayes factor BF_{ii_c} for H_i against H_{i_c}, and $BF_{ii'}$ for two competing hypotheses H_i and $H_{i'}$ can also be derived. Noting that the proportions of prior and posterior distributions in agreement with H_{i_c} are $1 - c_i$ and $1 - f_i$, respectively, it follows that

$$BF_{ii_c} = \frac{f_i}{c_i} \Big/ \frac{1 - f_i}{1 - c_i}. \tag{13}$$

Analogously, $BF_{ii'}$ can be obtained by

$$BF_{ii'} = BF_{iu}/BF_{i'u} = \frac{f_i}{c_i} \Big/ \frac{f_{i'}}{c_{i'}}. \tag{14}$$

Furthermore, an accessible manner for comparing a set of hypotheses is to transform Bayes factors into posterior model probabilities (PMPs). The PMPs are a representation

of the support in the data for each hypothesis on a scale between 0 and 1. Assuming equal prior probabilities for the hypotheses, we obtain PMPs for all the competing hypotheses excluding H_u using [5, p. 52]

$$PMP_i = \frac{BF_{iu}}{\sum_i BF_{iu}} \quad \text{for } i = 1, \ldots, I_N, \tag{15}$$

where I_N denotes the number of competing hypotheses. The execution of our programme renders both Bayes factors (10) and PMPs (15). As was shown in (10), the Bayes factor for H_i against H_u depends on the complexity and fit for which the prior and posterior distributions of $\boldsymbol{\theta}$ under H_u need to be specified, respectively. The specification of prior and posterior distributions will be introduced in the next section.

4. Prior and posterior distributions

4.1. Prior specification

The specification of prior distributions is an important step in Bayesian hypothesis testing. As can be seen from Equation (11), only a proper prior of $\boldsymbol{\theta}$ for the unconstrained hypothesis needs to be specified; the priors under the order constrained hypotheses automatically follow from this prior by truncating the unconstrained prior in the respective order constrained subspaces. The unconstrained prior that is proposed in this paper is partly based on the fractional Bayes factor of O'Hagan [22] which is known to be invariant for linear transformations. In the fractional Bayes factor a prior is implicitly constructed using a small fraction of the information in the data. A key property of the resulting automatic prior is that it has the same covariance structure as the covariance structure in the data [16].

In SEM the covariance structure in the data of the parameters of interest is contained in the estimated covariance matrix of the (standardized) target parameters, denoted by $\hat{\boldsymbol{\Sigma}}_\theta$. This covariance matrix can be obtained by standard SEM software packages such as `lavaan` [17]. Following the idea of a data-based covariance structure, as in the fractional Bayes factor, the following unconstrained normal prior will be used for the target parameters

$$\pi_u^*(\boldsymbol{\theta}) = N(\mathbf{0}, \omega\hat{\boldsymbol{\Sigma}}_\theta), \tag{16}$$

where ω controls the amount of prior information (a small/large value for ω implies an informative/vague prior). To avoid the dependence of the (arbitrarily chosen) mean vector $\mathbf{0}$, we let ω go to ∞. Although extremely vague priors are not recommended when testing hypotheses with equality constraints due to Lindley-Bartlett's paradox ([23]; Bartlett, 1957), such priors can be used for testing order constrained hypotheses [6].

To illustrate that the prior probability that the constraints hold is invariant for linear transformations, let us consider the following order constrained hypothesis $H_2 : \theta_1 > \theta_2 > \theta_3$ with restriction matrix

$$\boldsymbol{R}_2 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$$

for the repeated measures data $\boldsymbol{y}_i = (y_{i1}, y_{i2}, y_{i3})^{\mathrm{T}} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_y)$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^{\mathrm{T}}$ is a mean vector and $\boldsymbol{\Sigma}_y$ is a covariance matrix. Now let us consider a data set where the three

measurements are independent, e.g. $\hat{\mathbf{\Sigma}}_\theta = \mathbf{I}_3$, resulting in an unconstrained prior of the form $N(\mathbf{0}, \omega\mathbf{I}_3)$. In this case the prior probability, which reflects the relative complexity of H_2 relative to H_u, is equal to $Pr(\theta_1 > \theta_2 > \theta_3 | H_u) = \frac{1}{6}$, for any choice of $\omega > 0$.

Now we consider a one-to-one transformation of the data according to $\mathbf{z}_i = (y_{i1} - y_{i2}, y_{i2} - y_{i3}, y_{i3})^{\mathrm{T}} = \mathbf{L}\mathbf{y}_i$, with

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

We shall write $\mathbf{z}_i \sim N(\boldsymbol{\gamma}, \mathbf{\Sigma}_z)$, where $\boldsymbol{\gamma} = \mathbf{L}\boldsymbol{\theta} = (\theta_1 - \theta_2, \theta_2 - \theta_3, \theta_3)^{\mathrm{T}}$ and $\mathbf{\Sigma}_z = \mathbf{L}\mathbf{\Sigma}_y\mathbf{L}^{\mathrm{T}}$. Thus, the equivalent constrained hypothesis in the new parameterization corresponds to $H_2 : \gamma_1 > 0, \gamma_2 > 0$. Consequently, the estimated covariance matrix is now

$$\hat{\mathbf{\Sigma}}_\gamma = \mathbf{L}\hat{\mathbf{\Sigma}}_\theta\mathbf{L}^{\mathrm{T}} = \mathbf{L}\mathbf{L}^{\mathrm{T}} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

This results in an unconstrained prior for the target parameters of $(\gamma_1, \gamma_2)^{\mathrm{T}} \sim N\left(\mathbf{0}, \omega \left[\begin{smallmatrix} 2 & -1 \\ -1 & 2 \end{smallmatrix}\right]^{-1}\right)$. The prior probability now remains unchanged because $Pr(\gamma_1 > 0, \gamma_2 > 0 \,|\, H_u) = \frac{1}{6}$, for any choice of $\omega > 0$.

Theorem 4.1 provides a general proof of this invariance of the prior probability that the constraints of H_i hold with respect to the mean parameters in multivariate data.

Theorem 4.1: *The complexity of $H_i : \boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i$ when using $\pi_u^*(\boldsymbol{\theta})$ is invariant for linear one-to-one transformation of the multivariate data $\boldsymbol{y} \sim N(\boldsymbol{\theta}, \mathbf{\Sigma}_y)$.*

Proof: For the multivariate data, the covariance matrix of $\boldsymbol{\theta}$ is approximated by $\hat{\mathbf{\Sigma}}_\theta = \boldsymbol{S}_Y/n$, where $\boldsymbol{S}_Y = (\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^{\mathrm{T}})$ with $\bar{\boldsymbol{y}}$ being the sample means of $\boldsymbol{Y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$. Following (16) the prior distribution for $\boldsymbol{\theta}$ is $\pi_u^*(\boldsymbol{\theta}) = N(0, (\omega/n)\boldsymbol{S}_Y)$.

Consider a linear one-to-one transformation $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{z} \sim N(\boldsymbol{\gamma}, \mathbf{\Sigma}_z)$, where $\boldsymbol{L}$ is a $J \times J$ full rank matrix, and $\boldsymbol{\gamma} = \boldsymbol{L}\boldsymbol{\theta}$ and $\mathbf{\Sigma}_z = \boldsymbol{L}\mathbf{\Sigma}_y\boldsymbol{L}^{\mathrm{T}}$. After linear transformation, similarly, the covariance matrix of $\boldsymbol{\gamma}$ is approximated by $\hat{\mathbf{\Sigma}}_\gamma = \boldsymbol{S}_Z/n$, where $\boldsymbol{S}_Z = (\boldsymbol{Z} - \mathbf{1}\bar{\boldsymbol{z}}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{Z} - \mathbf{1}\bar{\boldsymbol{z}}^{\mathrm{T}})$ with $\bar{\boldsymbol{z}}$ being the sample means of $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$. Note that $\boldsymbol{S}_Z = \boldsymbol{L}(\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{Y} - \mathbf{1}\bar{\boldsymbol{y}}^{\mathrm{T}})\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{L}\boldsymbol{S}_Y\boldsymbol{L}^{\mathrm{T}}$ which implies $\hat{\mathbf{\Sigma}}_\gamma = \boldsymbol{L}\hat{\mathbf{\Sigma}}_\theta\boldsymbol{L}^{\mathrm{T}}$, then the prior distribution for $\boldsymbol{\gamma}$ becomes $\pi_u^*(\boldsymbol{\gamma}) = N(0, (\omega/n)\boldsymbol{L}\boldsymbol{S}_Y\boldsymbol{L}^{\mathrm{T}})$

Let $\boldsymbol{\beta}_1 = \boldsymbol{R}_i\boldsymbol{\theta} - \boldsymbol{r}_i$ and $\boldsymbol{\beta}_* = \boldsymbol{R}_i\boldsymbol{L}^{-1}\boldsymbol{\gamma} - \boldsymbol{r}_i$ with

$$\pi_u^*(\boldsymbol{\beta}_1) = N\left(0, \frac{\omega}{n}\boldsymbol{R}_i\boldsymbol{S}_Y\boldsymbol{R}_i^{\mathrm{T}}\right), \tag{17}$$

and

$$\pi_u^*(\boldsymbol{\beta}_*) = N\left(0, \frac{\omega}{n}\boldsymbol{R}_i\boldsymbol{L}^{-1}\boldsymbol{S}_Z(\boldsymbol{R}_i\boldsymbol{L}^{-1})^{\mathrm{T}}\right) = N\left(0, \frac{\omega}{n}\boldsymbol{R}_i\boldsymbol{S}_Y\boldsymbol{R}_i^{\mathrm{T}}\right) \tag{18}$$

then we have

$$\begin{aligned} P(\boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i \,|\, \pi_u^*(\boldsymbol{\theta})) &= P(\boldsymbol{\beta}_1 > 0 \,|\, \pi_u^*(\boldsymbol{\beta}_1)) = P(\boldsymbol{\beta}_* > 0 \,|\, \pi_u^*(\boldsymbol{\beta}_*)) \\ &= P(\boldsymbol{R}_i\boldsymbol{L}^{-1}\boldsymbol{\gamma} > \boldsymbol{r}_i \,|\, \pi_u^*(\boldsymbol{\gamma})) \end{aligned} \tag{19}$$

which manifests that the complexity is invariant. ■

Therefore the prior distribution in (16) is used for Bayes factor computation between order constrained hypotheses in SEM. Next, the posterior distribution is specified to obtain the relative fit (12).

4.2. Normal approximations to posterior distributions

In order to compute Bayes factors for order constrained hypotheses in SEM models, the asymptotic normality of the posterior distribution is used based on Laplace's method [[24], [25, pp. 101–107]]. As elaborated in the beginning of this section, the posterior distribution only depends on the density of the data $f(\boldsymbol{X} \mid \boldsymbol{\theta}, \boldsymbol{\zeta})$ when using the vague prior in (16) while letting $\omega \to \infty$. Subsequently the posterior distribution can be approximated by:

$$\pi_u(\boldsymbol{\theta} \mid \boldsymbol{X}) \approx N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_\theta), \tag{20}$$

where $\hat{\boldsymbol{\theta}}$ denotes the estimates of the target parameters, and $\hat{\boldsymbol{\Sigma}}_\theta$ is their covariance matrix. Both of them can be obtained in `lavaan` using estimation methods, such as least square estimation and maximum likelihood estimation [17]. Furthermore, to obtain standardized $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Sigma}}_\theta$ `lavaan` provides approaches to standardize the observed variables and to directly standardize the target parameters. The performance of these two approaches of standardization was discussed in [14], which showed that the variances of standardized parameters obtained using two approaches are different, whereas the resulting Bayes factors are similar. Now that the prior and posterior distributions have been specified, the Bayes factor can be obtained using (10). In the following section an efficient algorithm is described for the computation of the prior and posterior probability that the order constraints hold under H_u, which are key ingredients of the computation of the Bayes factor.

The normal approximation is widely used in hypothesis testing and model selection. Examples include Akaike's information criterion (AIC; [26]), Bayesian information criterion (BIC; [27]), and Wald's test [28]. In a way, the Bayes factor based on the approximated normal posterior (20) is similar as the BIC. The BIC is a large sample approximation of the marginal likelihood, whereas the proposed Bayes factor is a large sample approximation of a specific expression of the Bayes factor for an order constrained hypothesis against an unconstrained hypothesis. Both methods rely on a minimally informative prior and a large sample approximation of the posterior. An important difference is however that the proposed Bayes factor is suitable for evaluating hypotheses with order constraints while the BIC is not. To achieve this the proposed Bayes factor also needs the estimated Fisher information covariance matrix to approximate the posterior probability that the order constraints hold.

5. An efficient algorithm for Bayes factor computation

As was elaborated in Section 3, the Bayes factor is a ratio of the posterior probability that the order constraints of H_i hold under H_u, denoted by the relative fit f_i, and the prior probability that the order constraints of H_i hold under H_u, denoted by the relative complexity c_i. Because both the prior distribution $\pi_u^*(\boldsymbol{\theta}) = N(0, \omega\hat{\boldsymbol{\Sigma}}_\theta)$ and the posterior distribution $\pi_u(\boldsymbol{\theta} \mid \boldsymbol{X}) \approx N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_\theta)$ are normal distributions, for notational convenience each of them

can be denoted by

$$p(\boldsymbol{\theta}) = N(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta). \tag{21}$$

Thus, the complexity and fit can be represented by the following probability

$$P(H_i) = P(\boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i) = \int_{\boldsymbol{R}_i\boldsymbol{\theta}>\boldsymbol{r}_i} p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}. \tag{22}$$

This probability can be estimated by sampling from the prior or posterior distribution using the Gibbs sampler [25].

Before presenting the core algorithm of the Gibbs sampler, we shall present two pre-steps of the sampling procedure which can efficiently reduce the computing time. First, the Bayes factor is decomposed in Section 5.1 such that less iterations of the Gibbs sampler are needed to accurately estimate the complexity and fit. Second, the target parameters are transformed in Section 5.2 such that in each iteration of the Gibbs sampler less time is needed. Thereafter, Section 5.3 introduces the constrained Gibbs sampling procedure based on the transformed parameters. After obtaining the samples of transformed parameters, decomposed complexities and fits can be estimated using two methods proposed in Section 5.4. Furthermore, the sample size of the Gibbs sampler for accurate estimation of the complexity and fit is discussed in Section 5.5. Section 5.6 summarizes the constrained Gibbs sampling procedure by which we estimate the complexity and fit, and thus the Bayes factor.

5.1. Decomposition of the Bayes factor

When hypothesis H_i is formulated using a relatively large number of order constraints, accurately estimating the complexity and fit can be computationally intensive. For example, Gu et al. [14] showed that the complexity of $H_1 : \theta_1 > \cdots > \theta_{10}$ under prior

$$\pi_u^*(\boldsymbol{\theta}) = N(\mathbf{0}, \omega\mathbf{I})$$

is $c_1 = 1/J! = 1/10!$, that is, a very small value with the need of more than 20 million Gibbs sampler draws [5, p. 207] to ensure the deviation of the estimate is almost never over 10%. Directly estimating this complexity may not be feasible or extremely time-consuming. This conclusion also applies to the estimation of the complexity under $\pi_u^*(\boldsymbol{\theta})$ and the fit, because the accuracy of the estimation only depends on the size of complexity or fit and the number of Gibbs sampler draws. Consequently, when computing the Bayes factor for hypotheses with relatively large K, a decomposition of the Bayes factor is needed [29]:

$$BF_{iu} = BF_{i_1,u} \times BF_{i_2,i_1} \times \cdots \times BF_{i_K,i_{K-1}}, \tag{23}$$

where $i_k,\ k = 1, \ldots, K$ denotes a hypothesis using the constraints in the first k rows of $\boldsymbol{R}_i$. More specifically, $BF_{i_k,i_{k-1}}$ is defined by:

$$BF_{i_k,i_{k-1}} = \frac{f_{i_k,i_{k-1}}}{c_{i_k,i_{k-1}}}. \tag{24}$$

Let H_{i_k} denote the hypothesis using constraints in the first k rows of $\boldsymbol{R}_i$, then $c_{i_k,i_{k-1}}$ and $f_{i_k,i_{k-1}}$ denote the probabilities of prior and posterior distributions in agreement with H_{i_k}

conditional on $H_{i_{k-1}}$, respectively. Then, the complexity and fit can be expressed by

$$c_i = \prod_{k=1}^{K} c_{i_k,i_{k-1}} \quad \text{and} \quad f_i = \prod_{k=1}^{K} f_{i_k,i_{k-1}}. \tag{25}$$

Let

$$P(H_{i_k} \mid H_{i_{k-1}}) = P(\boldsymbol{R}_{i_k}\boldsymbol{\theta} > \boldsymbol{r}_{i_k} \mid \boldsymbol{R}_{i_1}\boldsymbol{\theta} > \boldsymbol{r}_{i_1}, \ldots, \boldsymbol{R}_{i_{k-1}}\boldsymbol{\theta} > \boldsymbol{r}_{i_{k-1}}) \tag{26}$$

denote either $c_{i_k,i_{k-1}}$ or $f_{i_k,i_{k-1}}$, then the probability (22) for c_i and f_i becomes

$$P(H_i) = P(H_{i_1}) \times P(H_{i_2}|H_{i_1}) \times \cdots \times P(H_{i_K}|H_{i_{K-1}}). \tag{27}$$

Because each of the probabilities in (27) is larger than $P(H_i)$ especially when K is large, accurately estimating $c_{i_k,i_{k-1}}$ or $f_{i_k,i_{k-1}}$ requires much less draws from the Gibbs sampler compared to directly estimating c_i or f_i. Although every probability in (27) needs to be estimated, the total sample size for decomposed c_i or f_i is still less than that without decomposition because the sample size for accurate estimation increases dramatically as K increases. This will be illustrated in Section 5.5. Before introducing the method for the computation of the probability (26), we transform the target parameters such that the order constrained hypothesis has a simple form, which will be elaborated in the next section.

5.2. Transformation of target parameters

This section simplifies the form of the hypothesis H_i using parameter transformation $\boldsymbol{\beta} = \boldsymbol{R}_i\boldsymbol{\theta} - \boldsymbol{r}_i$ such that $H_i : \boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i$ becomes $H_i : \boldsymbol{\beta} > 0$ and the decomposed complexity or fit shown in (26) becomes

$$P(H_{i_k} \mid H_{i_{k-1}}) = P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0). \tag{28}$$

This transformation was also used in [30]. It has three benefits in terms of the efficiency of estimating the decomposed complexity and fit. First, the subset of vector $\boldsymbol{\beta}$ that needs to be sampled has a length that is less than or equal to J (the length of $\boldsymbol{\theta}$). Take hypothesis $H_1 : \theta_1 > \theta_2 > \theta_3$ for example. The transformation $(\beta_1, \beta_2)^{\mathrm{T}} = (\theta_1 - \theta_2, \theta_2 - \theta_3)^{\mathrm{T}}$ leads to $H_1 : \beta_1 > 0, \beta_2 > 0$. Therefore, we only need to sample $\boldsymbol{\beta}$ with a length of 2. Although for another example $H_2 : \theta_1 > 0, \theta_2 > 0, \theta_1 > \theta_2$ the length of $\boldsymbol{\beta}$, where $(\beta_1, \beta_2, \beta_3)^{\mathrm{T}} = (\theta_1, \theta_2, \theta_1 - \theta_2)^{\mathrm{T}}$, is larger than the length of $\boldsymbol{\theta}$, only a subset $(\beta_1, \beta_2)^{\mathrm{T}}$ needs to be sampled because $\beta_3 = \beta_1 - \beta_2$. This issue will be further explained in the following paragraph. Second, it is more straightforward to define the conditional probability in (28) than in (26), because each β has a lower bound of 0 if it is constrained, whereas if θ is constrained, a lower and upper bound has to be determined which will take much effort especially when K is relatively large. It will be shown in Section 5.3 how the constrained $\boldsymbol{\beta}$ can be sampled. Third, the conditional probability $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ can analytically be determined, which will be further discussed in Section 5.4.

Since $\boldsymbol{\theta}$ has a multivariate normal distribution (21), after the linear transformation, $\boldsymbol{\beta}$ also has a multivariate normal distribution $p(\boldsymbol{\beta}) = N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, where $\boldsymbol{\mu}_\beta = \boldsymbol{R}_i\boldsymbol{\mu}_\theta - \boldsymbol{r}_i$ and $\boldsymbol{\Sigma}_\beta = \boldsymbol{R}_i\boldsymbol{\Sigma}_\theta\boldsymbol{R}_i^{\mathrm{T}}$. It should be noted that if $\boldsymbol{R}_i$ is of full row rank, then the elements of $\boldsymbol{\beta}$ is

linearly independent, otherwise the elements of $\boldsymbol{\beta}$ are not independent. Take, for example, hypothesis

$$H_3: \begin{array}{l}\theta_1 > \theta_3 \\ \theta_1 > \theta_4 \\ \theta_2 > \theta_3 \\ \theta_2 > \theta_4\end{array} \quad \text{with } [\boldsymbol{R}_3 \mid \boldsymbol{r}_3] = \left(\begin{array}{cccc:c} 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \end{array}\right). \tag{29}$$

The matrix $\boldsymbol{R}_3$ has a rank of 3 and the transformation

$$\boldsymbol{\beta} = \begin{pmatrix}\beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4\end{pmatrix} = \boldsymbol{R}_3\boldsymbol{\theta} - \boldsymbol{r}_3 = \begin{pmatrix}\theta_1 - \theta_3 \\ \theta_1 - \theta_4 \\ \theta_2 - \theta_3 \\ \theta_2 - \theta_4\end{pmatrix} \tag{30}$$

implies that $\beta_4 = -\beta_1 + \beta_2 + \beta_3$. Without loss of generality, we suppose the rank of $\boldsymbol{R}_i$ is M and let

$$\boldsymbol{\beta} = (\bar{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}) = (\bar{\beta}_1, \ldots, \bar{\beta}_M, \tilde{\beta}_{M+1}, \ldots, \tilde{\beta}_K), \tag{31}$$

where $\bar{\boldsymbol{\beta}}$ contains M independent elements of $\boldsymbol{\beta}$, and $\tilde{\boldsymbol{\beta}}$ is a linear combination of the elements of $\bar{\boldsymbol{\beta}}$. This implies that we only need to sample $\bar{\boldsymbol{\beta}}$ from its distribution. The distribution of $\bar{\boldsymbol{\beta}}$ is $p(\bar{\boldsymbol{\beta}}) = N(\boldsymbol{\mu}_{\bar{\beta}}, \boldsymbol{\Sigma}_{\bar{\beta}})$ with $\boldsymbol{\mu}_{\bar{\beta}} = \bar{\boldsymbol{R}}_i\boldsymbol{\mu}_\theta - \bar{\boldsymbol{r}}_i$ and $\boldsymbol{\Sigma}_{\bar{\beta}} = \bar{\boldsymbol{R}}_i\boldsymbol{\Sigma}_\theta\bar{\boldsymbol{R}}_i^{\mathrm{T}}$, where $\bar{\boldsymbol{R}}_i$ is a full row rank matrix that consists of M rows of $\boldsymbol{R}_i$ and $\bar{\boldsymbol{r}}_i$ is the corresponding constant vector. Although $\bar{\boldsymbol{R}}_i$ may not be unique, any set of linearly independent M rows of $\boldsymbol{R}_i$ can be chosen because the order of constraints does not affect the evaluation of the hypothesis.

The specification of $\bar{\boldsymbol{R}}_i$, $\bar{\boldsymbol{r}}_i$, and the linear combination of $\bar{\boldsymbol{\beta}}$ that renders $\tilde{\boldsymbol{\beta}}$ can be achieved using elementary row operations (Gaussian elimination) for the matrix $\boldsymbol{R}_i$. The procedure is implemented in `R` package `Bain`. Details are given as follows:

(1) Set an identity matrix $\boldsymbol{C}$ with a rank of $\max\{K, J\}$. Initialize $\boldsymbol{A} = \boldsymbol{R}_i$, $M = K$ and $\boldsymbol{d} = (1, 2, \ldots, K)$ to record the swap of constraints in $\boldsymbol{R}_i$.
(2) Repeat step (i), (ii) and (iii) for $k = 1, \ldots, K$.
   (i) If $\boldsymbol{A}_{k,k} = 0$ and $\boldsymbol{A}_{k',k} \neq 0$ where $k' > k$, then swap the kth row with the k'th row in $\boldsymbol{A}$ and $\boldsymbol{C}$, and swap d_k and $d_{k'}$ in $\boldsymbol{d}$.
   (ii) If $\boldsymbol{A}_{k,k} \neq 0$ after step (i), then let $\boldsymbol{A}_{k,j} = \boldsymbol{A}_{k,j}/\boldsymbol{A}_{k,k}$ and $\boldsymbol{C}_{k,j} = \boldsymbol{C}_{k,j}/\boldsymbol{C}_{k,k}$ for $j = 1, \ldots, J$.
   (iii) Let $\boldsymbol{A}_{k',j} = \boldsymbol{A}_{k',j} - \boldsymbol{A}_{k,j}\boldsymbol{A}_{k',k}$ and $\boldsymbol{C}_{k',j} = \boldsymbol{C}_{k',j} - \boldsymbol{C}_{k,j}\boldsymbol{C}_{k',k}$ for all $k' \neq k$ and $j = 1, \ldots, J$.
(3) For $k = 1, \ldots, K$, if $\sum_{j=1}^{J} |\boldsymbol{A}_{k,j}| = 0$ then $M = M-1$.
(4) For $k = 1, \ldots, K$, if $\sum_{j=1}^{J} |\boldsymbol{A}_{k,j}| = 0$ and $\sum_{j=1}^{J} |\boldsymbol{A}_{k',j}| \neq 0$ where $k' > k$, then swap the kth row with the k'th row in $\boldsymbol{A}$ and $\boldsymbol{C}$, and swap d_k and $d_{k'}$ in $\boldsymbol{d}$.
(5) Let $\boldsymbol{R}_i = (\boldsymbol{R}_{i,d_1}, \ldots, \boldsymbol{R}_{i,d_K})^{\mathrm{T}}$ and $\boldsymbol{r}_i = (r_{d_1}, \ldots, r_{d_K})$, where $\boldsymbol{R}_{i,d_k}$ denotes the d_kth row of $\boldsymbol{R}_i$. Then let $\boldsymbol{\beta} = \boldsymbol{R}_i\boldsymbol{\theta} > \boldsymbol{r}_i$ in which $\bar{\boldsymbol{\beta}}$ corresponds to the first M elements in $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$ corresponds to the remaining part.

After conducting this procedure, we obtain the rank of $\boldsymbol{R}_i$, i.e. M, and $[\bar{\boldsymbol{R}}_i|\bar{\boldsymbol{r}}_i]$ which contains the first M rows of $[\boldsymbol{R}_i|\boldsymbol{r}_i]$. Furthermore, the dependence in $\boldsymbol{\beta}$ can be expressed by

$$\begin{aligned} C_{M+1,d_1} \cdot \beta_1 + \cdots + C_{M+1,d_K} \cdot \beta_K &= r_{d_{M+1}}, \\ &\vdots \\ C_{K,d_1} \cdot \beta_1 + \cdots + C_{K,d_K} \cdot \beta_K &= r_{d_K}. \end{aligned} \tag{32}$$

For example, for the hypothesis H_3 shown in (29), executing the procedure above renders

$$\begin{aligned} [\boldsymbol{A} \mid \boldsymbol{C}] &= \left(\begin{array}{cccc:cccc} 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \end{array}\right) \\ &\rightarrow \left(\begin{array}{cccc:cccc} 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{array}\right) \end{aligned} \tag{33}$$

and $\boldsymbol{d} = (1, 3, 2, 4)$ which means the second and third rows have been swapped. Since there are three non-zero rows in $\boldsymbol{A}$ after Gaussian elimination, the rank of $\boldsymbol{R}_i$ is $M = 3$ and the first three rows of $\boldsymbol{R}_i$ are independent because they correspond to the non-zero rows. Furthermore, according to (32) the last row of $\boldsymbol{C}$ after Gaussian elimination indicates $\beta_1 - \beta_2 - \beta_3 + \beta_4 = 0$, i.e. $\beta_4 = -\beta_1 + \beta_2 + \beta_3$.

After the transformation of target parameters, the probability $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ from Equation (28) can be estimated using the constrained Gibbs sampler. This will be discussed in the next section.

5.3. Constrained Gibbs sampler

The constrained Gibbs sampler is applied to estimate each decomposed complexity and fit. The basic principle of the Gibbs sampler is to sequentially generate a sample for each β conditionally on the current values of all the others. As was elaborated before, only $\bar{\boldsymbol{\beta}}$ needs to be sampled, and the sample of $\tilde{\boldsymbol{\beta}}$ can be computed using the sample of $\bar{\boldsymbol{\beta}}$. Since $\bar{\boldsymbol{\beta}}$ is normally distributed, the conditional distribution of any parameter of $\bar{\boldsymbol{\beta}}$ given the remaining parameters is also normal. In each iteration, $\bar{\beta}_k^t$, where t denotes the iteration index of the Gibbs sampler and $k = 1, \ldots, M$, can be sampled from the following conditional distribution

$$p(\bar{\beta}_k^t \mid \bar{\beta}_{l\neq k}^t) = N\left(\mu_{\bar{\beta}_k} + \sum_{l=1}^{k-1} b_{kl}(\bar{\beta}_l^t - \mu_{\bar{\beta}_l}) + \sum_{l=k+1}^{M} b_{kl}(\bar{\beta}_l^{t-1} - \mu_{\bar{\beta}_l}), [(\boldsymbol{\Sigma}_{\bar{\beta}}^{-1})_{kk}]^{-1}\right), \tag{34}$$

where $\mu_{\bar{\beta}_k}$ is the mean of $\bar{\beta}_k$ in this full conditional distribution, b_{kl} is the element at the kth row and lth column in the matrix $\boldsymbol{B}_{M\times M} = \boldsymbol{I} - [\text{diag}(\boldsymbol{\Sigma}_{\bar{\beta}}^{-1})]^{-1}\boldsymbol{\Sigma}_{\bar{\beta}}^{-1}$ with $\boldsymbol{\Sigma}_{\bar{\beta}}$ being the covariance matrix of $\bar{\boldsymbol{\beta}}$ and $\boldsymbol{I}$ being a $M \times M$ identity matrix, and $(\boldsymbol{\Sigma}_{\bar{\beta}}^{-1})_{kk}$ is the element at

the kth row and kth column in $\boldsymbol{\Sigma}_{\bar{\beta}}^{-1}$. The derivation of equation (34) can be found in [25, p. 579].

The estimation of probability $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ requires a sample of $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_M)$ from the prior or posterior distribution that is in agreement with the first $k-1$ constraints $\beta_1 > 0, \ldots, \beta_{k-1} > 0$. Using the current value of $\boldsymbol{\beta}$ and the linear restriction if R_i is not of full row rank, a lower bound L and a upper bound U of $\bar{\boldsymbol{\beta}}$ can be specified. More specifically, if $k \leq M+1$ then $(\bar{\beta}_1, \ldots, \bar{\beta}_k)$ are sampled with a lower bound of $L=0$ and no upper bound, and other βs are not constrained. If $k > M+1$, all $\bar{\boldsymbol{\beta}}$ have a lower bound of $L=0$, and $(\tilde{\beta}_{M+1} > 0, \ldots, \tilde{\beta}_{k-1} > 0)$ will be used to define a further lower bound and a upper bound of $\bar{\boldsymbol{\beta}}$ based on their dependence. Using inverse probability sampling [31], it is straightforward to obtain a sample from truncated normal distribution (34) constrained in (L, U) according to the following two steps.

(i) Randomly generate a number ν via a uniform distribution on the interval [0,1].
(ii) Compute $\bar{\beta}_k = \Phi_{\bar{\beta}_k}^{-1}[\Phi_{\bar{\beta}_k}(L) + \nu(\Phi_{\bar{\beta}_k}(U) - \Phi_{\bar{\beta}_k}(L))]$, where $\Phi_{\bar{\beta}_k}$ is the cumulative distribution function of (34) and $\Phi_{\bar{\beta}_k}^{-1}$ is the inverse cumulative distribution function.

Running the Gibbs sampler for $t = 1, \ldots, T$ iterations renders a sample of each component of $\bar{\boldsymbol{\beta}} = (\bar{\beta}_1, \ldots, \bar{\beta}_M)$. As elaborated in Section 5.2, $\tilde{\boldsymbol{\beta}}$ is linearly dependent on $\bar{\boldsymbol{\beta}}$. Thus, we can also obtain a sample of $\tilde{\boldsymbol{\beta}}$ using the sample of $\bar{\boldsymbol{\beta}}$ and Equation (32).

The choice of burn-in period and the check of convergence are important steps in the Gibbs sampler. In our method, however, we specify the prior distribution and approximate the posterior distribution with a multivariate normal distribution. Therefore, convergence is not an issue because the sample from multivariate normal distribution converges very fast even if the initial value is far away from the prior or posterior mode. This is explicitly illustrated in [14], which applies the constrained Gibbs sampler to multivariate normal distributions as well. In addition, Gu et al. [14] also shows that within a burn-in period of 100 iterations the effect of the initial values is eliminated and the sample converges to the desired distribution. Thus, we discard the first 100 iterations as a burn-in phase of the Gibbs sampler. In the next section, two methods for estimating the decomposed complexity and fit are presented based on the samples of $\boldsymbol{\beta}$ obtained in this section.

5.4. Two methods for estimating complexity and fit

In this section, we propose two approaches to estimate the probability (28) after obtaining the samples of $\boldsymbol{\beta}$ of size T from either prior or posterior distribution. A straightforward manner is counting the number of samples in agreement with $\beta_k > 0$:

$$P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0) = T^{-1} \sum_{t=1}^{\mathrm{T}} I(\beta_k^t > 0 \mid \beta_1^t > 0, \ldots, \beta_{k-1}^t > 0), \quad (35)$$

where $I(\cdot)$ denotes the indicator function which is 1 if the argument is true and 0 otherwise.

Particularly for estimating this probability with respect to the first M decomposed constraints $\bar{\boldsymbol{\beta}} > 0$, we adopt a more efficient approach inspired by Gelfand and Smith [31] and used in [30,32]. The principle of this method is that the density of the univariate β_k can

be approximated by the average of its full conditional density constructed using the current sample of all the other βs. This implies the probability $P(\beta_k > 0)$ given the density of β_k can be approximated by the average of $P(\beta_k > 0)$ given the conditional density based on different samples. Consequently, using the constrained samples for $\beta_1, \ldots, \beta_{k-1}$ in the conditional density, we obtain

$$P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0) = T^{-1} \sum_{t=1}^{\mathrm{T}} P(\beta_k > 0 \mid \beta_1^t > 0, \ldots, \beta_{k-1}^t > 0, \beta_{k+1}^t, \ldots, \beta_K^t). \quad (36)$$

This probability can easily be computed because the conditional distribution (34) of β_k is a univariate normal distribution that is easily integrated for $\beta_k > 0$.

It should be emphasised that this method is not applicable for estimating decomposed complexities or fits for which $k > M$, because $\tilde{\beta}_k$ for $k > M$ is a linear combination of $\bar{\beta}_1, \ldots, \bar{\beta}_M$, which means $\tilde{\beta}_k$ is a point given $\bar{\beta}_1, \ldots, \bar{\beta}_M$. Therefore in this case Equation (35) will be used. Despite of this limitation, the new method (36) is still attractive because it increases the accuracy of the estimation for a give sample size of the Gibbs sampler. This will be elaborated in the next paragraph. This implies that fewer iterations of the Gibbs sampler are needed to obtain an acceptable accuracy. Consequently, for estimating the decomposed complexities and fits in our programme, the new method (36) is used when $k \leq M$, whereas the approach shown in (35) is used when $k > M$.

To investigate the performance of the two methods, we shall consider a series of hypotheses $H_1 : \theta_1 > \ldots > \theta_J$ for $J = 3, \ldots, 5$ and estimate the complexities under $\pi_u^*(\theta) = N(\mathbf{0}, \omega \hat{\mathbf{\Sigma}}_\theta)$, where $\mathbf{0}$ is a zero vector with a length of J, $\hat{\mathbf{\Sigma}}_\theta = \boldsymbol{I}$ is a $J \times J$ identity matrix, and $\omega \to \infty$. The true value of c_1 with respect to prior $\pi_u^*(\theta)$ in these hypotheses is known as $c_1^{True} = 1/J!$. We estimate the complexities of H_1 1000 times using each method when the sample size of the Gibbs sampler is $T = 50$, 500, and 5000. This results in $c_{11}^{(s)}$ and $c_{12}^{(s)}$ based on methods (35) and (36), respectively, where $s = 1, \ldots, 1000$. Thereafter, we compute the mean squared relative error (MSRE), $MSRE_1 = \frac{1}{1000} \sum_{s=1}^{1000} ((c_1^{True} - c_{11}^{(s)})/c_1^{True})^2$ and $MSRE_2 = \frac{1}{1000} \sum_{s=1}^{1000} ((c_1^{True} - c_{12}^{(s)})/c_1^{True})^2$, to measure the accuracy of the estimation using methods (35) and (36), respectively.

Table 1 displays the MSREs of the estimate for c_1. As can be seen in Table 1, the MSREs from method (36) MSRE_2 are much smaller than that from method (35) MSRE_1. This implies that method (36) needs a smaller sample size of the Gibbs sampler to attain the same accuracy. Furthermore, it can be seen that the MSREs decreases as sample size increases, and small complexity $c_i = 8.33\mathrm{E}-3$ needs more sample size than large complexity $c_i = 0.166$ to obtain the same magnitude of MSREs. This implies we can determine sample size

Table 1. MSRE of estimate using two methods.

True	$c_i = 0.166$ ($J = 3$)		$c_i = 4.17\mathrm{E}-2$ ($J = 4$)		$c_i = 8.33\mathrm{E}-3$ ($J = 5$)	
	MSRE_1	MSRE_2	MSRE_1	MSRE_2	MSRE_1	MSRE_2
$T = 50$	7.76E−2	8.37E−3	0.140	3.36E−2	0.272	9.64E−2
$T = 500$	9.25E−3	7.62E−4	1.61E−2	3.34E−3	2.44E−2	8.96E−3
$T = 5000$	5.28E−4	7.78E−5	1.49E−3	3.38E−4	2.46E−3	9.15E−4

T for both methods (35) and (36) based on the acceptable estimation accuracy and the size of the probability under estimation. This will be discussed in the next section.

5.5. Sample size determination for the Gibbs sampler

This section discusses the sample size T of the Gibbs sampler needed to accurately estimate $P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$, which has a true value P^{True}. As stated earlier, this probability is estimated using method (35) if $k > M$, and method (36) if $k \leq M$. For method (35), Hoijtink [5] proposes a rule to determine the sample size T_1 needed to accurately estimate the complexity or fit, which is shown in the top panel of Table 2. The criterion is that the 95% central credibility interval for the estimate has lower and upper bounds that are less than 10% different from the true value. The first row in Table 2 displays the true probabilities P^{True} that needs to be estimated. In addition, L-95% and U-95% demonstrate the lower and upper bounds of the 95% central credibility interval when using the corresponding T_1 above.

For method (36), we present a new rule to determine the sample size T_2 based on a more strict accuracy criterion, that is, the differences between both L-95% and U-95%, and P^{True} are less than 5%. We let $N(\mu_{\beta_k}, \sigma^2_{\beta_k})$ denote the distribution of β_k in $P(\beta_k | \beta_1 > 0, \ldots, \beta_{k-1} > 0)$, where μ_{β_k} is the mean and $\sigma^2_{\beta_k}$ is the variance. Then Equation (36) becomes

$$\begin{aligned} P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0) &= P(\beta_k > 0 \mid \beta_k \sim N(\mu_{\beta_k}, \sigma^2_{\beta_k})) \\ &= P(\beta_k > 0 \mid \beta_k \sim N(\hat{\lambda}_k, 1)), \end{aligned} \tag{37}$$

where $\hat{\lambda}_k = \mu_{\beta_k}/\sigma_{\beta_k}$ is the standardized population mean of β_k. The principle of the sample size determination for method (36) is based on two facts. First, in the Gibbs sampler, we obtain T_2 samples of β_k from $N(\mu_{\beta_k}, \sigma^2_{\beta_k})$ or standardized β_k from $N(\hat{\lambda}_k, 1)$. This implies that the distribution of the standardized sample mean of β_k, denoted by λ_k, is $N(\hat{\lambda}_k, 1/T_2)$. Second, the probability $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ is a one-to-one correspondence function of $\hat{\lambda}_k$. For example, if $\hat{\lambda}_k = 0$, we obtain a probability of 1/2, and conversely if the true value of the probability is 1/6, we would expect a $\hat{\lambda}_k$ of -0.97. These enable us to determine the sample size T_2 needed to accurately estimate $P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ given a true value P^{True} using the following steps.

(1) Compute $\hat{\lambda}_k$ such that $P(\beta_k > 0 \mid \beta_k \sim N(\hat{\lambda}_k, 1)) = P^{True}$, and initialize $T_2 = 1000$.
(2) Sample λ_k 10,000 times from $N(\hat{\lambda}_k, 1/T_2)$, and then obtain 10,000 estimates of $P(\beta_k > 0 \mid \beta_k \sim N(\hat{\lambda}_k, 1))$.

Table 2. Gibbs sample size determination.

P^{True}	0.166	4.17E−2	8.33E−3	1.39E−3	1.98E−4	2.48E−5
T_1	3,000	9,600	120,000	360,000	2,520,000	20,160,000
L-95%	0.154	3.8E−2	7.8E−3	1.27E−3	1.82E−4	2.3E−5
U-95%	0.180	4.6E−2	8.9E−3	1.52E−3	2.17E−4	2.7E−5
T_2	4,000	8,000	12,000	18,000	25,000	32,000
L-95%	0.159	3.97E−2	7.93E−3	1.32E−3	1.89E−4	2.37E−5
U-95%	0.175	4.36E−2	8.74E−3	1.46E−3	2.08E−4	2.60E−5

(3) Using 10,000 estimates of $P(\beta_k > 0 \mid \beta_k \sim N(\hat{\lambda}_k, 1))$, compute their 95% central credibility interval (L, U).
(4) If either $|L - P^{True}|/P^{True} > 5\%$ or $|U - P^{True}|/P^{True} > 5\%$, then $T_2 = T_2 + 1000$ and go to Step 2.

The bottom panel of Table 2 displays the sample size T_2 and the resulting L-95% and U-95% from the procedure above given corresponding P^{True}.

In `Bain`, Table 2 is adopted to determine the sample size T_1 and T_2 of the Gibbs sampler for estimating each decomposed complexity and fit based on methods (35) and (36). Because T_1 or T_2 is large enough to accurately estimate the corresponding P^{True} in the first row of Table 2, it will also be sufficient to estimate a probability that is larger than this P^{True}. We estimate $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ with a starting sample size $T_1 = 3000$ if $k > M$ or $T_2 = 4000$ if $k \leq M$, and gradually reset T_1 or T_2 based on Table 2 until the estimate of the complexity or fit is larger than the corresponding P^{True}. Note that if the estimate is still less than 2.48E-5 when using the corresponding T_1 or T_2, we specify $T_1 = 100,000,000$ or $T_2 = 100,000$.

5.6. Summary of the computation of the Bayes factor

This section summarizes the computation of the Bayes factor for H_i against H_u, which is a ratio of the fit and complexity. The following steps describe how our programme computes the complexity and fit, and therefore the Bayes factor.

1. Transform $\boldsymbol{\theta}$ into $\boldsymbol{\beta}$ using the procedure shown in Section 5.2. Then, we obtain $(\bar{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}})$ and M the rank of $\boldsymbol{R}_i$.
2. Repeat Step (1), ..., (6) for $k = 1, \ldots, K$ to estimate each $P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ for the decomposed complexity $c_{i_k, i_{k-1}}$ and fit $f_{i_k, i_{k-1}}$.
   1 Initialize the sample size of the Gibbs sampler as $T_2 = 4000$ if $k \leq M$ and $T_1 = 3000$ if $k > M$, and initialize $\boldsymbol{\beta} = 0$.
   2 Repeat Step (a) or (b) for $t = 1, \ldots, T_2 + 100$ iterations if $k \leq M$ or for $t = 1, \ldots, T_1 + 100$ iterations if $k > M$, where 100 denotes the first 100 iterations, that is, a burn-in phase of the Gibbs sampler.
      a If $k \leq M + 1$, then define a boundary $(L, U) = (0, \infty)$ for $\bar{\beta}_1, \ldots, \bar{\beta}_{k-1}$ and no boundary for $\bar{\beta}_k, \ldots, \bar{\beta}_K$. Thereafter, sequentially generate a sample of $\bar{\boldsymbol{\beta}}^t$ from the truncated distribution of (34) as previously described in Step (i) and (ii) in Section 5.3.
      b If $k > M + 1$, then define a boundary (L, U) for $\bar{\beta}_1, \ldots, \bar{\beta}_M$ using the linear relation between the $\bar{\boldsymbol{\beta}} > 0$ and $\tilde{\boldsymbol{\beta}} > 0$. Thereafter, sequentially generate a sample of $\bar{\boldsymbol{\beta}}^t$ from the truncated distribution of (34) as previously described in Step (i) and (ii) in Section 5.3. Then a sample of $\tilde{\boldsymbol{\beta}}^t$ is obtained by means of its linear dependence on $\bar{\boldsymbol{\beta}}^t$.
   3 Discard all the iterations for which $t \leq 100$ to account the burn-in period as discussed in Section 5.3.
   4 If $k \leq M$, compute the probability $P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0) = T_2^{-1} \sum_{t=101}^{T_2+100} P(\beta_k > 0 \mid \beta_k \sim N(\mu_{\beta_k}^t, (\sigma_{\beta_k}^2)^t))$ using method (36) in Section 5.4.

5 If $k > M$, compute the probability $P(\beta_k > 0 \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0) = T_1^{-1} \sum_{t=101}^{T_1+100} I(\beta_k^t > 0 \mid \beta_1^t > 0, \ldots, \beta_{k-1}^t > 0)$ using method (35) in Section 5.4.

6 If $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ obtained in Step (4) or (5) is less than the reference value that corresponds to the current T_2 or T_1 in Table 2, respectively, then reset T_2 or T_1 using the value of the next column in the table and restart the procedure from Step (2). If not, the estimation of $P(\beta_k \mid \beta_1 > 0, \ldots, \beta_{k-1} > 0)$ is completed, which renders the decomposed complexity $c_{i_k,i_{k-1}}$ or fit $f_{i_k,i_{k-1}}$. This was elaborated in Section 5.5.

3. The complexity and fit can be computed by $c_i = \prod_{k=1}^{K} c_{i_k,i_{k-1}}$ and $f_i = \prod_{k=1}^{K} f_{i_k,i_{k-1}}$ shown in Section 5.1. Then, the Bayes factor for H_i against H_u is $BF_{iu} = f_i/c_i$.

6. Empirical applications in SEM

In this section, our procedure of evaluating order constrained hypotheses will be illustrated using two classic SEM applications. One example concerns confirmatory factor analysis (CFA), and the other example concerns multiple regression model.

6.1. Confirmatory factor analysis

In the first example, we reanalyze a dataset built into `lavaan` called HolzingerSwineford1939 [17]. This dataset is taken from the Holzinger and Swineford 1939 (H&S) study, which is a commonly used example in factor analysis. The raw dataset consists of scores of 301 seventh and eighth grade students from the Pasteur School ($n = 145$) and Grant-White School ($n = 156$) who participated in 26 psychological aptitude tests. In our example, only a subset with 9 variables of the complete data is extracted to measure 3 correlated latent variables, each with three indicators, i.e.

- a visual factor (ξ_1) is measured by visual perception (x_1), cubes (x_2) and lozenges (x_3).
- a textual factor (ξ_2) is measured by paragraph comprehension (x_4), sentence completion (x_5) and word meaning (x_6).
- a speed factor (ξ_3) is measured by addition (x_7), counting of dots (x_8) and discrimination of straight and curved capitals (x_9).

The descriptives for the observed variables are given in Table 3, whereas the relations between latent variables and their indicators are formulated in the next paragraph and expressed using path notation (without showing measurement errors) in Figure 1.

The confirmatory factor analysis model for the H&S data can be represented as:

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\epsilon}_x, \tag{38}$$

where $\boldsymbol{x} = (x_1, \ldots, x_9)^{\mathrm{T}}$ denotes observed variables, $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^{\mathrm{T}}$ denotes latent variables,

$$\boldsymbol{\Lambda}_x^{\mathrm{T}} = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \theta_4 & \theta_5 & \theta_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_7 & \theta_8 & \theta_9 \end{pmatrix} \tag{39}$$

Table 3. Descriptives for the variables in the confirmatory factor analysis.

Variable		Mean	S.D.
visual perception	x_1	4.94	1.17
cubes	x_2	6.09	1.18
lozenges	x_3	2.25	1.13
paragraph	x_4	3.06	1.16
sentence	x_5	4.34	1.29
word mean	x_6	2.19	1.10
addition	x_7	4.19	1.09
dots	x_8	5.53	1.01
straight curved	x_9	5.37	1.01

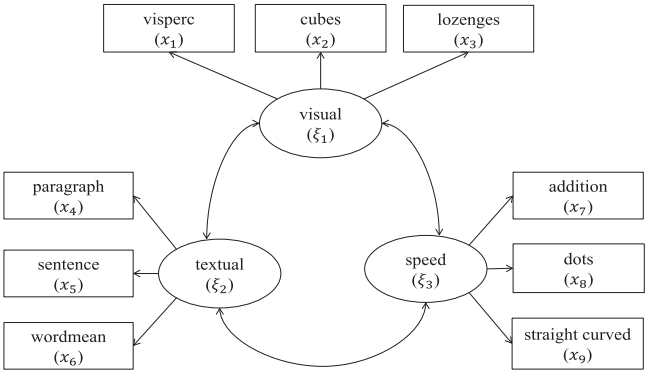

Figure 1. Confirmatory factor analysis.

is a matrix of factor loadings, and $\boldsymbol{\epsilon}_x$ is a 3×1 vector of measurement errors with $\boldsymbol{\epsilon}_x \sim N(0, \boldsymbol{\Psi}_{\epsilon_x})$ and $\boldsymbol{\Psi}_{\epsilon_x}$ being its covariance matrix. The covariance matrix of observed variables is given by:

$$\boldsymbol{\Sigma}_x = \boldsymbol{\Lambda}_x \boldsymbol{\Phi}_\xi \boldsymbol{\Lambda}_x^{\mathrm{T}} + \boldsymbol{\Psi}_{\epsilon_x}, \tag{40}$$

where the factor covariance matrix $\boldsymbol{\Phi}_\xi$ is a symmetric matrix:

$$\boldsymbol{\Phi}_\xi = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{12} & \phi_{22} & \phi_{23} \\ \phi_{13} & \phi_{23} & \phi_{33} \end{pmatrix}. \tag{41}$$

Because the confirmatory factor analysis model is a measurement model without a structural model, we can simply specify this model using `lavaan` syntax in `R` (see Appendix 1). To ensure that the target parameters are comparable, we standardize them all. As is elaborated in Appendix 1, `lavaan` provides both the standardized estimates and covariance matrix of target parameters. Recall that this is all the information that `Bain` needs to compute Bayes factors. Furthermore, in factor analysis models, indicators are required to both identify the model and set a metric for latent variables. This can be typically achieved either by standardizing the variances of latent variables or by constraining one factor loading per latent variable to 1. In this example, the former way is chose.

Factor loadings indicate the degree of correspondence between the factor and the indicator, with higher loadings making the indicator more representative of the factor.

Researchers might be interested in the issue which indicator plays the most important role in defining a factor. For instance, the first indicator of every factor may be expected to be strongest, which can be represented by the following hypothesis

$$H_1 : \begin{array}{c} \theta_1 > \{\theta_2, \theta_3\} \\ \theta_4 > \{\theta_5, \theta_6\}. \\ \theta_7 > \{\theta_8, \theta_9\} \end{array} \tag{42}$$

We can also test a hypothesis with respect to the structure of the correlations between the latent variables. For example, we can evaluate whether the correlation between visual and textual is larger than the correlation either between visual and speed or between textual and speed:

$$H_2 : \phi_{12} > \{\phi_{13}, \phi_{23}\}. \tag{43}$$

Using `R` package `Bain` (the user manual of `Bain` can be found in Appendix 2) to compute Bayes factors for H_1 against H_u or H_{1_c} renders $BF_{1u} = 0.076$ or $BF_{11_c} = 0.073$. For H_2 against H_u or H_{2_c}, `Bain` renders $BF_{2u} = 1.33$ or $BF_{22_c} = 1.59$. These results imply that hypothesis H_1 is not supported by the data, and the evidence from the data for H_2 is not convincing because BF_{2u} or BF_{22_c} is quite close to 1.

6.2. Multiple regression with latent variables

In a study reported by Warren et al. [33] (data available at https://informative-hypotheses.sites.uu.nl/software/bain), a sample of 98 managers of farmer cooperatives was selected with the objective of studying managerial behaviour. They postulated that a latent variable manager performance (η) was predicted by three correlated latent variables, i.e. knowledge (ξ_1), orientation (ξ_2) and satisfaction (ξ_3), and an observed variable training (x_4). The latent variables η, ξ_1, ξ_2, and ξ_3 were measured based on qualitative and quantitative answers to identical questionnaires collected from a random sample of managers in farmer cooperatives. These variables are assumed to be measured with error, and the errors of measurement were computed using the split halves procedure [33] for all variables subject to measurement error:

- η is measured by y_1 and y_2,
- ξ_1 is measured by x_{11} and x_{12},
- ξ_2 is measured by x_{21} and x_{22},
- ξ_3 is measured by x_{31} and x_{32}.

The observed variables are described in Table 4 and the graphical specification of this structural equation model is found in Figure 2.

As can be seen from Figure 2, the relations of the variables can be represented by a multiple regression model with η, ξ_1, ξ_2, and ξ_3 that are latent. The measurement model is given by

$$\begin{aligned} \boldsymbol{y} &= \Lambda_y \eta + \boldsymbol{\epsilon}_y \\ \boldsymbol{x} &= \Lambda_x \boldsymbol{\xi} + \boldsymbol{\epsilon}_x, \end{aligned} \tag{44}$$

Table 4. Descriptives for the variables in the multiple regression model.

Variable	Mean	S.D.
y_1	1.06	0.16
y_2	1.05	0.15
x_{11}	1.43	0.30
x_{12}	1.33	0.24
x_{21}	2.84	0.43
x_{22}	2.91	0.38
x_{31}	2.54	0.34
x_{32}	2.47	0.32
x_4	2.12	0.31

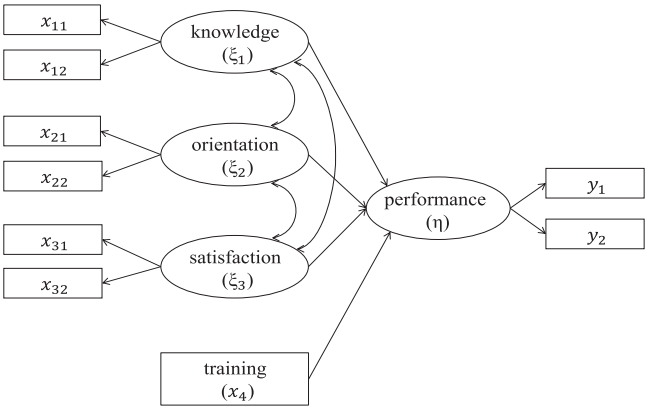

Figure 2. Multiple regression with latent variables.

where $\boldsymbol{x} = (x_{11}, x_{12}, x_{21}, x_{22}, x_{31}, x_{32})^{\mathrm{T}}$ denotes observed variables, and η and $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^{\mathrm{T}}$ are latent variables. For the structural model, we have

$$\eta = \theta_0 + \theta_1\xi_1 + \theta_2\xi_2 + \theta_3\xi_3 + \theta_4 x_4 + \delta, \tag{45}$$

where θ_0 is the intercept, θ_1, θ_2, θ_3, and θ_4 are regression coefficients, and $\delta \sim N(0, \sigma^2)$ is the residual. This regression model is analyzed in `lavaan` (see Appendix 1). We standardize the coefficients to make them comparable. Using the standardized estimates and covariance matrix of these coefficients from `lavaan`, `Bain` can compute Bayes factors.

The hypothesis we evaluated is based on the results obtained by Warren et al. [33] It states that knowledge is the strongest predictor followed by orientation, training and satisfaction. The resulting hypothesis is

$$H_3 : \theta_1 > \theta_2 > \theta_4 > \theta_3. \tag{46}$$

This hypothesis can be compared to, for example, knowledge is stronger than orientation followed by satisfaction and training:

$$H_4 : \theta_1 > \theta_2 > \theta_3 > \theta_4, \tag{47}$$

Table 5. Bayes factors and PMPs of H_3, H_4 and H_5.

	BF_{ic}	PMPs
H_3	11.90	0.785
H_4	2.676	0.214
H_5	0.010	0.001

and training is stronger than satisfaction followed by orientation and knowledge:

$$H_5 : \theta_4 > \theta_3 > \theta_2 > \theta_1. \tag{48}$$

The results of the evaluation of these three hypotheses using `Bain` are displayed in Table 5 (the user manual of `Bain` can be found in Appendix 2). As can be seen, there is evidence in favour of H_3, no convincing evidence for H_4, and evidence against H_5. Furthermore, it can be seen from the PMPs introduced in (15) that H_3 receives the largest support from the data.

7. Discussion

Order constrained hypotheses provide a representation of a researcher's theory with respect to the relations between the parameters of interest in SEM models. We developed a Bayes factor that can evaluate these hypotheses in a direct manner. A very vague prior was proposed that incorporates the covariance structure of the target parameters in the data. A proof was given that the prior probability that the order constraints hold, a key ingredient of the Bayes factor when testing order constrained hypotheses, was invariant for linear transformations of the data.

The multivariate normal prior that is used to compute the prior probability can be applied to order constrained testing problems where parameters have symmetric prior distributions such as regression coefficients, group means, and factor loadings. Even in the case of non-symmetric prior distributions, the procedure will be accurate in most cases. For example, when testing a specific ordering of J variances using identical inverse gamma priors, the prior probability of this specific ordering will be equal to $1/J!$ which is identical to when computing the probability using independent normal priors. The method could break down in asymmetric cases where the boundary value does not lie in the middle of a parameter space. For instance when testing $0 \leq \theta < .2$ versus $.2 \leq \theta \leq 1$, where θ is the probability of a success in a binomial experiment, and a uniform prior is specified on θ, the prior probabilities of θ falling in these two intervals could be different than when computing the probability using a normal distribution on θ. Extending the methodology for such asymmetric testing problems would be an interesting topic to explore for future research.

Furthermore, a new algorithm was developed to ensure fast computation to ensure general utilization of the methodology by applied researchers. The methodology was implemented in the R package `Bain` which only needs the estimates and covariance matrix of target parameters (which can be obtained from the free `R`-package `lavaan`), and one or more restriction matrices representing a researcher's expectations. The output from `Bain` consists of Bayes factors and posterior probabilities for the hypotheses. These can be used which provide a direct answer about the relative evidence in the data between the hypotheses under investigation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Herbert Hoijtink is supported by the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). Joris Mulder was supported by a Vidi Grant of the Netherlands Organization of Scientific Research (NWO grant number 452-17-006).

References

[1] Cohen J. The earth is round, $p < .05$. American Psychologist. 1994;49:997–1003.
[2] Royall R. Statistical evidence: a likelihood paradigm. New York, NY: Chapman and Hall/CRC; 1997.
[3] Silvapulle MJ, Sen PK. Constrained statistical inference; order, inequality, and shape constraints. New York, NY: Wiley; 2005.
[4] van de Schoot R, Hoijtink H, Deković M. Testing inequality constrained hypotheses in SEM models. Struct Equ Model. 2010;17:443–463.
[5] Hoijtink H. Informative hypotheses: theory and practice for behavioral and social scientists. Boca Raton, FL: Chapman and Hall/CRC; 2012.
[6] Klugkist I, Laudy O, Hoijtink H. Inequality constrained analysis of variance: a Baysian approach. Psychol Methods. 2005;10:477–493.
[7] van de Schoot R., Hoijtink H, Hallquist MN, et al. Bayesian evaluation of inequality constrained hypotheses in SEM models using Mplus. Struct Equ Model. 2012;19:1–17.
[8] Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90:773–795.
[9] Wagenmakers EJ. A practical solution to the pervasive problems of *p* values. Psychon Bull Rev. 2007;14:779–804.
[10] Kuiper RM, Hoijtink H. Comparisons of means using exploratory and confirmatory approaches. Psychol Methods. 2010;15:69–86.
[11] Kuiper RM, Klugkist I, Hoijtink H. A `Fortran 90` program for confirmatory analysis of variance. J Stat Softw. 2010;34(8):1–31.
[12] Mulder J, Hoijtink H, Klugkist I. Equality and inequality constrained multivariate linear models: objective model selection using constrained posterior priors. J Stat Plan Inference. 2010;140:887–906.
[13] Mulder J, Hoijtink H, de Leeuw C. BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. J Stat Softw. 2012;46(2):1–39.
[14] Gu X, Mulder J, Deković M, et al. Bayesian evaluation of inequality constrained hypotheses. Psychol Methods. 2014;19(4):511–527.
[15] Mulder J. Bayes factors for testing inequality constrained hypotheses: issues with prior specification. British J Math Statist Psych. 2014;67:153–171.
[16] Mulder J. Prior adjusted default Bayes factors for testing inequality constrained hypotheses. Comput Stat Data Anal. 2014;71:448–463.
[17] Rosseel Y. lavaan: an `R` package for structural equation modeling. J Stat Softw. 2012;48(2):1–36.
[18] Jöreskog K, Sörbom D. Advances in factor analysis and structural equation models. Cambridge, MA: Abt books; 1979.
[19] Bollen KA. Structural equations with latent variables. New York, NY: John Wiley and Sons; 1989.
[20] Jeffreys H. Theory of probability. 3rd ed. Oxford: Oxford University Press; 1961.
[21] Klugkist I, Hoijtink H. The Bayes factor for inequality and about equality constrained models. Comput Stat Data Anal. 2007;51:6367–6379.

[22] O'Hagan A. Fractional Bayes factors for model comparison. J R Stat Soc Ser B. 1995;57:99–138.
[23] Lindley D. A statistical paradox. Biometrika. 1957;44:187–192.
[24] DiCiccio TJ, Kass RE, Raftery A, et al. Computing Bayes factors by combining simulation and asymptotic approximations. J Am Stat Assoc. 1997;92:903–915.
[25] Gelman A, Carlin JB, Stern HS, et al. Bayesian data analysis. 2nd ed. Boca Raton, FL: Chapman and Hall/CRC; 2004.
[26] Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csáki F, editors. Proc. 2nd Int. Symp. Information Theory. Budapest: Akademiai kiado; 1973. p. 267–281.
[27] Schwarz G. Estimating the dimension of a model. Ann Statist. 1978;6(2):461–464.
[28] Gourieroux C, Holly A, Monfort A. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. Econometrica. 1982;50:63–80.
[29] Klugkist I, Laudy O, Hoijtink H. Bayesian evaluation of inequality and equality constrained hypotheses for contingency tables. Psychol Methods. 2010;15:281–299.
[30] Mulder J. Bayes factors for testing order-constrained hypotheses on correlations. J Math Psychol. 2016;72:104–115.
[31] Gelfand AE, Smith AFM, Lee T. Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. J Am Stat Assoc. 1992;87:523–532.
[32] Morey RD, Rouder JN, Pratte MS, et al. Using MCMC chain outputs to efficiently estimate Bayes factors. J Math Psychol. 2011;55:368–378.
[33] Warren RD, White JK, Fuller WA. An errors-in-variables analysis of managerial role performance. J Am Stat Assoc. 1974;69:886–893.

Appendices

Appendix 1. Estimates and covariance matrix obtained using lavaan

`Bain` uses the estimates and covariance matrix of target parameters to compute Bayes factors. These can be obtained from the `R` package `lavaan` [17]. This appendix illustrates how to obtain the estimates and covariance matrix of target parameters using the two examples discussed in Section 6.

First of all, researchers need to install the version 0.5-18 or higher version of `lavaan` by starting `R` and typing `install.packages(''lavaan'')`. Note that `R` should be upgraded to `R.3.5.0` or a higher version. The user manual of the latest version of `lavaan` can be found at `https://CRAN.R-project.org/package=lavaan`.

The following `R` syntax renders the estimates and covariance matrix for the CFA model presented in Section 6.1.

```
# Load lavaan package.
  library(lavaan)
# Specify the CFA model.
  CFA.model <- 'visual  =~ x1 + x2 + x3
                textual =~ x4 + x5 + x6
                speed   =~ x7 + x8 + x9'
  fit<-cfa(CFA.model,data=HolzingerSwineford1939,std.lv = TRUE)

# Obtain standardized estimates of parameters
  standardizedSolution(fit)
# Obtain standardized covariance matrix of parameters.
  Sigma <- lavInspect(fit, "vcov.std.all")
  Sigma[1:9,1:9]   # For target parameters in (42)
  Sigma[19:21,19:21] # For target parameters in (43)
```

The output of `standardizedSolution(fit, ci = FALSE)` for the CFA model is

```
      lhs op     rhs est.std    se      z pvalue
1   visual =~      x1   0.772 0.055 14.041      0
2   visual =~      x2   0.424 0.060  7.105      0
3   visual =~      x3   0.581 0.055 10.539      0
4  textual =~      x4   0.852 0.023 37.776      0
5  textual =~      x5   0.855 0.022 38.273      0
6  textual =~      x6   0.838 0.023 35.881      0
7    speed =~      x7   0.570 0.053 10.714      0
8    speed =~      x8   0.723 0.051 14.309      0
9    speed =~      x9   0.665 0.051 13.015      0
⋮
22  visual ~~ textual   0.459 0.064  7.189      0
23  visual ~~   speed   0.471 0.073  6.461      0
24 textual ~~   speed   0.283 0.069  4.117      0
```

Note that the label `visual =~ x1` denotes the factor loading θ_1 relating x_1 to ξ_1 and the label `visual ~~ textual` denotes the covariance ϕ_{12} between ξ_1 and ξ_2. We only show the results for nine factor loadings used in (42) and three covariances used in (43). The standardized estimates of the target parameters are given in the column under `est.std`. For example, the estimate of θ_4 is 0.852 in the row of `textual =~ x4` , and the estimate of ϕ_{23} is 0.283 in the row of `textual ~~ speed`.

The output of `Sigma` contains the standardized covariance matrix of the target parameters. We only show the covariance matrix `Sigma[19:21,19:21]` of ϕ_{12}, ϕ_{13}, and ϕ_{23}:

```
               visual~~textual  visual~~speed  textual~~speed
 visual~~textual  0.0040678110   0.0007276616     0.001156340
 visual~~  speed  0.0007276616   0.0053037342     0.001480068
textual~~  speed  0.0011563398   0.0014800678     0.004723718
```

The following R syntax renders the estimates and covariance matrix for the regression model in Section 6.2.

```
# Load lavaan package.
  library(lavaan)
# Set R working director where the data is saved.
  setwd("C:/Example2")
# Read data "performance.csv".
  performance<-read.csv("performance.csv")

# Specify the regression model.
  perform.model<-'
              # measurement model
                 kno =~ x11 + x12
                 ori =~ x21 + x22
                 sat =~ x31 + x32
                 per =~ y1 + y2
              # regressions
                 per ~ kno + ori + sat + tra'
  fit<-sem(perform.model,data=performance,std.lv = TRUE)

# Obtain standardized estimates and covariance matrix
  standardizedSolution(fit)
  Sigma <- lavInspect(fit, "vcov.std.all")
  Sigma[9:12,9:12] # For target parameters in (46), (47), (48)
```

The output of `standardizedSolution(fit, ci = FALSE)` for the regression model is

```
   lhs op rhs est.std    se     z pvalue
⋮
9  per  ~ kno   0.478 0.161 2.960  0.003
10 per  ~ ori   0.336 0.165 2.030  0.042
11 per  ~ sat   0.151 0.105 1.440  0.150
12 per  ~ tra   0.286 0.084 3.403  0.001
⋮
```

Note that the label `per ~ kno` denotes the coefficient θ_1 which relates η to ξ_1 in the regression model (45). We only show the results for the four regression coefficients used in (46), (47), and (48). The standardized estimates of the target parameters are given in the column under `est.std`. For example, the estimate of θ_1 is 0.478 in the row of `per ~ kno`, and the estimate of θ_4 is 0.286 in the row of `per ~ tra`.

The output of `Sigma[9:12,9:12]` renders the standardized covariance matrix of $\theta_1, \ldots, \theta_4$:

```
           per~kno       per~ori       per~sat       per~tra
per~kno  0.026034895 -0.0223249106 -0.0050273595 -0.0011610045
per~ori -0.022324911  0.0273346337  0.0043904540 -0.0007619234
per~sat -0.005027359  0.0043904540  0.0110250662 -0.0002713825
per~tra -0.001161004 -0.0007619234 -0.0002713825  0.0070519650
```

The standardized estimates and covariance matrix of target parameters obtained in `lavaan` can be used as input for `Bain`. This will be shown in the user manual in Appendix 2.

Appendix 2. User manual of Bain

`Bain` is an `R` package developed for the evaluation of order constrained hypotheses using the algorithm presented in the paper. It can be downloaded at https://informative-hypotheses.sites.uu.nl/software/bain/. Windows, Mac, and Linux versions are offered, respectively, by downloaded files `Bain_xxx.zip`, `Bain_xxx.tgz`, and `Bain_xxx.tar.gz`, where `xxx` denotes package version. After downloading the package, for example, windows users can install `Bain` in `R` by

```
install.packages("···/Bain_xxx.zip", repos = NULL)
```

This appendix provides a brief user manual of `Bain`. The detailed manual can be found on the website. The core function of `Bain` package is

```
Bain(estimate, Sigma, grouppara = 0, jointpara = 0, n,
ERr = NULL, IRr = NULL,..., seed = 100, print = TRUE).
```

The input arguments contain the estimates and covariance matrix of target parameter, number of target parameters, sample size, and the restriction matrix for each hypothesis under consideration. The output of `Bain` are the Bayes factor and PMP for each hypothesis. We will use the example from Section 6.2 to illustrate the use of `Bain`.

The estimates and covariance matrix of target parameters can be obtained using `R` package `lavaan` as shown in Appendix 1. For example, from the output of `lavaan` for the regression model, we observe in `R` syntax:

```
# estimates
estimate<-c(0.478, 0.336, 0.151, 0.286)

# Covariance matrix
Sigma<-matrix(c(0.026, -0.022, -0.005, -0.001,
                -0.022,  0.027,  0.004, -0.001,
                -0.005,  0.004,  0.011, -0.000,
                -0.001, -0.001, -0.000,  0.007), nrow = 4)
```

In addition, the sample size is `n = 98`, and the number of target parameters is `jointpara = 4`. Argument `grouppara` indicates the number of group specific parameters, which is zero because in the regression model there is no group specific parameter. Furthermore, argument `IRr` specifies order constrained hypotheses, while argument `ERr` for equality constraints will not be used since this paper only deals with order constrained hypotheses. The paragraph below demonstrates how `IRr` can be constructed to represent order constrained hypotheses.

As was shown in Section 2.2, an order constrained hypothesis H_i can be formulated by $\boldsymbol{R}_i\theta > \boldsymbol{r}_i$. Each constraint $\boldsymbol{R}_{ik}\boldsymbol{\theta} > r_{ik}$ for $k = 1, \ldots, K$ in the hypothesis can be written as $R_{ik1}\theta_1 + \ldots + R_{ikJ}\theta_J > r_{ik}$, where K and J are numbers of constraints and parameters in H_i, respectively. Note that every parameter should be moved to the left hand side of the inequality sign '$>$', and the constant should be moved to the right hand. In the restriction matrix `IRr`, the constraint $\boldsymbol{R}_{ik}\boldsymbol{\theta} > r_{ik}$ can be expressed by the line

$R_{ik1}\ R_{ik2}\ \ldots\ R_{ikJ}\ \ r_{ik}$.

For example,

- $\theta_1 + \theta_2 + \theta_3 > 0$ corresponds to
  1 1 1 0
- $\theta_1 - 2\theta_2 + 3\theta_3 > 0.5$ corresponds to
  1 -2 3 0.5
- $\theta_1 - 2 > \theta_2 - \theta_3$ corresponds to
  1 -1 1 2
- $\theta_1 > \theta_2 > \theta_3$ corresponds to
  1 -1 0 0
  0 1 -1 0
- $\theta_1 - \theta_2 > \theta_3 - \theta_4 > \theta_5 - \theta_6$ corresponds to
  1 -1 -1 1 0 0 0
  0 0 1 -1 -1 1 0

Thus, in the regression model in Section 6.2, three competing order constrained hypotheses $H_3 : \theta_1 > \theta_2 > \theta_4 > \theta_1$, $H_4 : \theta_1 > \theta_2 > \theta_3 > \theta_4$, and $H_5 : \theta_4 > \theta_3 > \theta_2 > \theta_1$ can be represented in `R` script, respectively, by

```
# order constrained hypotheses
IRr1<-matrix(c(1, -1, 0, 0, 0,
               0, 1, 0, -1, 0,
               0, 0, -1, 1, 0), nrow = 3, byrow = TRUE)

IRr2<-matrix(c(1, -1, 0, 0, 0,
               0, 1, -1, 0, 0,
               0, 0, 1, -1, 0), nrow = 3, byrow = TRUE)

IRr3<-matrix(c(-1, 1, 0, 0, 0,
               0, -1, 1, 0, 0,
               0, 0, -1, 1, 0), nrow = 3, byrow = TRUE)

# no equality constrained hypotheses
ERr1<-ERr2<-ERr3<-NULL
```

Once the estimates and covariance matrix of parameters, number of parameters, sample size, and order constrained hypotheses are specified, running the following line in `R` renders test results for the example of regression model:

```
res<-Bain(estimate, Sigma, grouppara = 0, jointpara = 4, n = 98,
          ERr1, IRr1, ERr2, IRr2, ERr3, IRr3)
```

The output of `Bain` function is stored in a list which consists of `$testResult` and `$BFmatrix`. The `$testResult` reports fits, complexities, Bayes factors and PMPs of each hypothesis under consideration. The `$BFmatrix` reports Bayes factor matrix for competing hypotheses from which users can easily obtain Bayes factor $BF_{ii'}$ for one hypothesis against another. Take again the regression model for example, the output are

```
round(res$testResult,3)
      fit  complexity      BF   PMPa   PMPb
H1  0.217       0.023  11.902  0.786  0.726
H2  0.049       0.019   2.676  0.214  0.197
H3  0.000       0.019   0.010  0.001  0.001

round(res$BFmatrix,3)
      H1      H2       H3
H1 1.000  3.677  914.137
H2 0.272  1.000  248.614
H3 0.001  0.004    1.000
```

Note that in the first table `BF` displays Bayes factors of order constrained hypotheses against their complements. In addition, `PMPa` lists PMPs excluding the unconstrained hypothesis, whereas `PMPb` includes. In the second table, we can observe for example the Bayes factor for H_1 against H_2 is $BF_{12} = 3.677$, and the Bayes factor for H_2 against H_1 is $BF_{21} = 0.272$.