Bayesian Evaluation of Inequality-Constrained Hypotheses in SEM Models Using Mplus

Rens van de Schoot,1,2 Herbert Hoijtink,1 Michael N. Hallquist,3 and Paul A. Boelen1
1Utrecht University
2North-West University
3University of Pittsburgh

Researchers in the behavioral and social sciences often have expectations that can be expressed in the form of inequality constraints among the parameters of a structural equation model resulting in an informative hypothesis. The questions they would like an answer to are “Is the hypothesis Correct” or “Is the hypothesis incorrect?” We demonstrate a Bayesian approach to compare an inequality-constrained hypothesis with its complement in an SEM framework. The method is introduced and its utility is illustrated by means of an example. Furthermore, the influence of the specification of the prior distribution is examined. Finally, it is shown how the approach proposed can be implemented using Mplus.

Keywords: Bayes factor, informative hypothesis, Mplus, order restricted inference, structural equation modeling

Many researchers in the behavioral and social sciences have expectations that can be expressed in the form of inequality constraints among the parameters of a structural equation model. Based on previous research, literature reviews, and the current academic debate, they formulate order restrictions among, for example, regression coefficients. For example, in health psychology research, an important distinction is the distinctiveness of symptoms of complicated grief (CG) and uncomplicated grief (UG; Boelen & van den Bout, 2008). As we discuss in more details later, CG symptoms should be expected to be more strongly associated with symptoms of depression and anxiety than responses of UG. The questions they would like an answer to are “Is the hypothesis correct” or “Is the hypothesis incorrect?” The evaluation
of inequality-constrained hypotheses has been studied within the structural equation modeling (SEM) framework (Gonzalez & Griffin, 2001; Stoel, Galindo-Garre, Dolan, & Van den Wittenboer, 2006; Van de Schoot, Hoijtink, & Deković, 2010; Van de Schoot & Strohmeier, 2011) as well as outside the SEM framework (Barlow, Bartholomew, Bremner, & Brunk, 1972; Hoijtink, Klugkist, & Boelen, 2008; Klugkist, Laudy, & Hoijtink, 2005; Robertson, Wright, & Dykstra, 1988; Silvapulle & Sen, 2004; Van de Schoot, Hoijtink, et al., 2011; Van de Schoot, Mulder, et al., 2011). The procedures described in these studies tested an inequality-constrained hypothesis against the classical null hypothesis (i.e., “Nothing is going on”), against an unconstrained hypothesis (i.e., “Something is going on, but not the null hypothesis”), or against other inequality-constrained hypotheses (i.e., CG symptoms are expected to be more strongly associated with symptoms of depression than responses of UG, but the opposite holds for anxiety). For an overview of the literature see Van de Schoot, Romeijn, and Hoijtink (2011).

A comparison of an inequality-constrained hypothesis with its complement received hardly any attention in the literature, but only this latter comparison would provide an answer to the questions of interest: “Is the hypothesis correct” or “Is the hypothesis incorrect?” One exception is the study of van Rossum, Van de Schoot, & Hoijtink (2011; see also Hoijtink, 2012), in the context of analysis of variance. In this article we propose a method for the evaluation of an inequality-constrained hypothesis with respect to its complement using Bayesian statistics available in the software Mplus (Muthén & Muthén, 1998–2010).

In what follows we first introduce Bayesian SEM, followed by the formulation of inequality-constrained hypotheses in Bayesian SEM models, and the use of Bayes factors to evaluate such hypotheses. Then we discuss specification of the prior distribution for the evaluation of inequality-constrained hypotheses. Finally, we apply the proposed methodology to a real-life data set from clinical and health psychology.

**BAYESIAN SEM MODELS**

For an introduction to Bayesian estimation see Lynch (2007); for more technical details see Gelman, Carlin, Stern, and Rubin (2004); for Bayesian model selection see Hoijtink et al. (2008) and Hoijtink (2012); and for Bayesian SEM see Lee (2004). There is a variety of software available for Bayesian analyses: R packages (e.g., mcmc), WinBUGS, AMOS, OpenBUGS, MLwiN, and Mplus v6.x. For this article we use Mplus (1998–2010) because of its popularity among applied researchers. Bayesian estimation first became available in Mplus in version 6, which was released in the summer of 2010. For a technical implementation of Bayesian statistics in Mplus, see Asparouhov & Muthén (2010a, 2010b) and the website of Mplus at www.statmodel.com.

Let \( y_i = (y_1, \ldots, y_p) \) be a \( p \times 1 \) vector of observed variables for person \( i \), and let \( \omega_i = (\omega_1, \ldots, \omega_q) \) be a \( q \times 1 \) vector of latent variables for person \( i \). The measurement model is then given by

\[
y_i = \Lambda \omega_i + \epsilon_i,
\]

where \( \Lambda \) is a \( p \times q \) matrix of factor loadings, and \( \epsilon \) is a \( p \times 1 \) vector of measurement errors with \( \epsilon \sim N(0, \Psi_\epsilon) \), and \( \Psi_\epsilon \) being a diagonal matrix.
For the structural model, let \( \mathbf{\eta}_i = \{\eta_i, \xi_i\} \) with \( \eta = (\eta_1, \ldots, \eta_{q_1}) \) being a \( q_1 \times 1 \) vector of outcomes variables and \( \xi = (\xi_1, \ldots, \xi_{q_2}) \) being a \( q_2(= p - q_1) \times 1 \) vector of explanatory latent variables, then
\[
\eta_i = \mathbf{\Pi} \eta_i + \mathbf{\Gamma} \xi_i + \delta_i,
\]
where \( \mathbf{\Pi} \) and \( \mathbf{\Gamma} \) are matrices of unknown regression coefficients. Furthermore, \( \delta \sim N(0, \mathbf{\Psi}_\delta) \) and \( \xi \sim N(0, \Phi) \). It is assumed that \( \epsilon \) and \( \delta \) are independent. Equations 1 and 2 describe the general SEM framework.

Let \( \mathbf{\theta} = \{\mathbf{\theta}^1, \mathbf{\theta}^2\} \), which consist of unknown parameters with \( \mathbf{\theta}^1 = \text{vec}\{\mathbf{\Pi}, \mathbf{\Gamma}\} \) and \( \mathbf{\theta}^2 = \{\mathbf{\Lambda}, \mathbf{\Psi}_\epsilon, \mathbf{\Psi}_\delta, \Phi\} \), and let \( Y \) be the observed data set with sample size \( n \). In a Bayesian SEM model \( \mathbf{\theta} \) is considered to be random. The behavior of \( \mathbf{\theta} \) given \( Y \) in such a Bayesian model can be described by \( g(\mathbf{\theta}|Y) \propto h(\mathbf{\theta}) \times f(Y|\mathbf{\theta}) \), where \( f(Y|\mathbf{\theta}) \) is the likelihood function, \( h(\mathbf{\theta}) \) is the prior distribution, and \( g(\mathbf{\theta}|Y) \) is the posterior distribution.

Using the posterior distribution the mean, mode, or median of the desired parameters can be estimated. Bayesian estimation uses Markov chain Monte Carlo (MCMC) algorithms to obtain a sample from \( g(\mathbf{\theta}|Y) \). The idea behind MCMC is that the conditional distribution of one set of parameters given other sets can be used to make random draws of parameter values, ultimately resulting in an approximation of the joint distribution of all the parameters. The default MCMC sequence used in \textit{Mplus} is a Gibbs sampler where \( t \) iterations \( (t = 1, \ldots, T) \) are used to obtain new values for \( \mathbf{\theta} \) in each step drawing from a conditional posterior parameter distribution. The Gibbs sampler has three steps:

1. Sample \( \eta_i \) and \( \xi_i \) given \( Y, \mathbf{\Pi}, \mathbf{\Gamma}, \mathbf{\Lambda}, \Phi, \mathbf{\Psi}_\epsilon, \mathbf{\Psi}_\delta \).
2. Sample \( \mathbf{\Pi}, \mathbf{\Gamma}, \mathbf{\Lambda}, \Phi \) given \( Y, \eta_i, \xi_i, \mathbf{\Psi}_\epsilon, \mathbf{\Psi}_\delta \).
3. Sample \( \mathbf{\Psi}_\epsilon \) and \( \mathbf{\Psi}_\delta \) given \( Y, \eta_i, \xi_i, \mathbf{\Pi}, \mathbf{\Gamma}, \mathbf{\Lambda}, \Phi \).

These three conditional distributions have been derived in Lee (2004) and for the implementation in \textit{Mplus} see Asparouhov and Muthen (2010a). In \textit{Mplus} it is easy to sample from these conditional distributions. Typically, several MCMC chains are used (default in \textit{Mplus} is two chains) and some part of each chain is discarded as a burn-in phase (in \textit{Mplus} by default the first half). \textit{Mplus} automatically monitors convergence using the Gelman–Rubin convergence criterion, which considers within- and between-chain variability of the parameter estimates (Asparouhov & Muthen, 2010b; Gelman et al., 2004). In \textit{Mplus} it is possible to first estimate the SEM model using maximum likelihood estimation and use these parameter estimates as starting values in the Bayesian analysis.

The default prior specification in \textit{Mplus} is
\[
h(\mathbf{\theta}^1) = \prod_{k=1}^{\mathbf{K}} h(\mathbf{\theta}^1_k),
\]
where \( h(\mathbf{\theta}^1_k) = \mathcal{N}(\theta_0, \tau^2_0) \) with \( \theta_0 = 0 \) and \( \tau^2_0 = \infty \). In \textit{Mplus} this prior distribution is used for means and intercepts of observed and latent continuous variables, thresholds of observed categorical dependent variables, factor loadings, and regression coefficients. For the prior distribution of \( \mathbf{\theta}^2 \), \textit{Mplus} uses noninformative inverse Gamma distributions for variances.
and residual variances of observed and latent parameters, an inverse Wishart distribution if more than one latent variable is estimated, and a Dirichlet distribution for categorical latent variable parameters. Later it will be shown that the default prior distribution used by Mplus is suited for the evaluation of inequality-constrained hypotheses by means of the Bayes factor (which are elaborated in the next section). This is because the default prior distribution leads to an adequacy quantification of the complexity of inequality-constrained hypotheses (elaborated in the section after the next).

INEQUALITY-CONSTRAINED HYPOTHESES

In SEM, researchers typically have expectations about $\theta_1^1$ and not about $\theta_1^2$. Hence, we focus on inequality constraints imposed on the $\Pi$ or $\Gamma$ matrix. If $M$ is the number of inequality constraints imposed on $\theta_1^1$, and $K$ the number of parameters involved, then let $A$ be an $M \times K$ matrix of full rank with known constants. Now, an inequality-constrained hypothesis can be specified as follows:

$$H_i : A\theta_1 > d,$$

where $d$ is an $m \times 1$ vector of known constants.

Consider a hypothetical example and say, based on some theory, we expect the following ordering of four regression coefficients $H_i : \theta_1^1 > \theta_1^2 > \theta_1^3 > \theta_1^4$, then

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad \theta_1^1 = \begin{bmatrix} \theta_1^1 \\ \theta_1^2 \\ \theta_1^3 \\ \theta_1^4 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which renders

$$H_i = \begin{bmatrix} \theta_1^1 - \theta_1^2 > 0 \\ \theta_1^2 - \theta_1^3 > 0 \\ \theta_1^3 - \theta_1^4 > 0 \end{bmatrix},$$

which is equivalent to $H_i : \theta_1^1 > \theta_1^2 > \theta_1^3 > \theta_1^4$.

Another example is $H_i : \theta_1^2 - \theta_1^1 > \theta_1^3 - \theta_1^2 > 0.10$, where the difference between the first two regression coefficients is larger than the difference between the last two regression coefficients. Moreover, the difference between the coefficients is expected to be larger than 0.10. This hypothesis can be constructed using:

$$A = \begin{bmatrix} -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \quad \theta_1^1 = \begin{bmatrix} \theta_1^1 \\ \theta_1^2 \\ \theta_1^3 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0.10 \end{bmatrix}. $$
Note that the requirement that $A$ is of full rank excludes equality constraints like $|\theta_1^1 - \theta_2^1| < d$, which could be constructed using

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and

$$d = [-d, -d].$$

**BAYES FACTORS FOR THE EVALUATION OF $H_i$**

Typically applied researchers test the hypothesis

$$H_0 : \theta^1 = 0,$$

or

$$H_0 : \theta_1^1 = \theta_2^1 = \ldots = \theta_K^1,$$

for $K$ parameters against the alternative hypothesis

$$H_1 : \text{not } H_0.$$  

We argue, however, that researchers are not really interested in such a test (Hoijtink et al., 2008; Klugkist et al., 2005; Van de Schoot, Hoijtink, et al., 2011; Van de Schoot, Romeijn, & Hoijtink, 2011; Wagenmakers, 2007). Rather, they are interested in the evaluation of their expectations directly. These expectations are about the expected ordering of, for example, regression coefficients and are based on previous research or a literature study. Therefore we propose to evaluate the inequality-constrained hypothesis presented in Equation 4 instead of the null hypotheses in Equations 10 or 11.

The evaluation of $H_i$ versus either $H_0$ or $H_1$ using $p$ values has been described by, for example, Silvapulle and Sen (2004), Van de Schoot, Hoijtink, et al. 2011, and Van de Schoot, Mulder, et al. 2011). However, $H_i$ is contained in $H_1$, which might be considered logically inconsistent. Moreover, we argue that many applied researchers are interested in the questions “Is the hypothesis correct” or “Is the hypothesis incorrect?” Therefore we explore the comparison of $H_i$ versus its complement,

$$H_{-i} : \text{not } H_i.$$  

As was shown by Mulder et al. (2009) and Mulder, Hoijtink, and Klugkist (2010), the Bayes factor of $H_i$ versus an unconstrained hypothesis, $H_u$, can be written as

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

where $f_i$ can be interpreted as model fit, or the proportion of the posterior distribution of $H_u$ in agreement with the inequality-constrained hypothesis $H_i$. Furthermore, $c_i$ can be interpreted
as the complexity of $H_i$ or as the proportion of the prior distribution of $H_u$ in agreement with the constraints of $H_i$. Using $f_{-i} = 1 - f_i$ and $c_{-i} = 1 - c_i$,

$$BF_{-i} = \frac{f_{-i}}{c_{-i}} = \frac{1 - f_i}{1 - c_i}. \quad (15)$$

Combining Equations 14 and 15 leads to

$$BF_{-i} = \frac{BF_i}{BF_{-i}} = \frac{f_i/c_i}{(1 - f_i)/(1 - c_i)}. \quad (16)$$

The resulting Bayes factor can be interpreted as a relative measure of support for the research questions “Is the hypothesis correct” and “Is the hypothesis incorrect?” If $BF_{-i} > 1$, $H_i$ is more supported by the data than $H_{-i}$. If $BF_{-i} \approx 1$, neither of the two hypotheses is preferred by the data. For $BF_{-i} < 1$, $H_{-i}$ is more supported by the data than $H_i$. See Kass and Raftery (1995) for more information about the Bayes factor and its interpretation and for the interpretation in the context of inequality-constrained hypotheses, see Van de Schoot, Mulder, et al. (2011).

**COMPLEXITY**

In this section we focus on the prior distribution, $h(\theta^1)$ and we show that the default prior specification (Equation 3) that is used in Mplus is suitable for the computation of the Bayes factor (Equation 16). This is because this default specification results in an adequate quantification of the complexity of inequality-constrained hypotheses.

Hoijtink (2012) defined the complexity of inequality-constrained hypotheses using an argument based on equivalent sets of hypotheses. Consider, for example, the hypothesis $H_i \in \{ \theta_1^1 > \theta_2^1 > \theta_3^1 \}$. There are $3! = 6$ permutations of the $\theta^1$'s rendering hypotheses that have an equivalent structure. A logical implication is to consider these hypotheses to be of the same complexity. This leads to the following definition:

**Definition 1.** An equivalent set consists of equivalent hypotheses $H_{i1}, \ldots, H_{iZ}$ for which $H_{i1} \cup \ldots \cup H_{iZ}$ encompasses 100% of the parameter space.

With $3! = 6$ equivalent hypotheses, 16.6% of the parameter space is in agreement with each hypothesis, with $Z$ equivalent hypotheses the proportion of the parameter space in agreement with each is $1/Z$. This leads to the following definition of complexity:

**Definition 2.** The complexity of an inequality-constrained hypothesis is the proportion of the parameter space in agreement with $H_{iz}$.

Let $A_{m} = \{A_{m1}, \ldots, A_{mk}\}$ denote the $m$th row from $A$. The following restrictions on $A$ render a hypothesis that is a member of an equivalent set:

1. Each $A_{mk} \in \{-1, 0, 1\}$.
2. For $m = 1, \ldots, M$, $\sum_k A_{mk} = 0$. 


3. $A_1$ can be divided into $D$ subsets of the same size, such that for $m = 2, \ldots, M$, $A_m$ is a permutation of these subsets.

The number of elements in an equivalent set can be determined as follows:

1. Obtain all $D!$ permutations of the $D$ subsets in $\theta^1$; that is, divide $\theta^1_1, \ldots, \theta^1_K$ in the same subsets as $A_{11}, \ldots, A_{1K}$.
2. Denote the number of permutations for which $A^1 > 0$ is in agreement with $H_i$ by $B$.
3. Then $Z = D!/B$.

The following theorem shows that prior distributions of the form in Equation 16 render a complexity measure in agreement with Definitions 1 and 2 for hypotheses that belong to an equivalent set:

**Theorem 1** (for Proof see Appendix A)

Let

$$h(\theta^1, \theta^2) = h(\theta^1)h(\theta^2).$$

For prior distributions with the property that $h(\theta^1_1, \ldots, \theta^1_K) = h(\theta^1_{(1)}, \ldots, \theta^1_{(K)})$ for all permutations $(1), \ldots, (K)$ of $1, \ldots, K$, it holds that $c_i = 1/Z$ for each hypothesis in an equivalent set.

Note that the default prior used by Mplus (see Equation 3), is in agreement with the requirements specified in Theorem 1.

Not all inequality-constrained hypotheses belong to an equivalent set. Consider, for example,

$$0_1^1 - 0_2^1 > 0_3^1 - 0_4^1$$

$$H_{i1} :$$

$$0_1^1 > 0_2^1$$

$$0_3^1 > 0_4^1.$$  \hspace{1cm} (18)

Each element of $H_{i1}$ can be permuted in two ways; however, although all $2 \times 2 \times 2$ permutations cover 100% of the parameter space, not each permutation is equally complex. For example, 0% of the parameter space is in agreement with

$$0_1^1 - 0_2^1 > 0_3^1 - 0_4^1$$

$$H_{i2} :$$

$$0_2^1 > 0_1^1$$

$$0_3^1 > 0_4^1.$$  \hspace{1cm} (19)

Consider also $H_{i1} : 0_1^1 > 0, 0_2^1 > 0$. If the ordinate $(0,0)$ is treated as the natural midpoint of the parameter space, the permutation argument renders four hypotheses (another is $H_{i2} : 0_1^1 < 0, 0_2^1 < 0$) that are of equal complexity. Note that Jeffreys (1961) and Berger and Mortera
(2006) in the univariate counterpart of \( H_i \) use 0 as the mean of the prior distribution and thus as the natural midpoint of the parameter space.

It can be proven for hypotheses of the form Equation 4, and thus also for hypotheses belonging to equivalent sets, that the complexity of a hypothesis is independent of the mean and variance of the prior distribution in Mplus.

**Theorem 2 (for Proof see Appendix B)**

If \( h(\theta^1_k) = \mathcal{N}(\theta_0, \tau^2_0) \) for \( k = 1, \ldots, K \) and \( A \) is of full rank, \( c_i \) is independent of \( \theta_0 \) for \( \tau^2_0 \to \infty \).

An implication of Theorem 2 for hypotheses of the form \( H_i : A\theta^1 > d \) can be illustrated using the hypothesis \( H_i : \theta^*_1 > \theta^*_2 > \theta^*_3 \) where \( \theta^*_1 = \theta_1, \theta^*_2 = \theta_2 + 1 \) and \( \theta^*_3 = \theta_3 + 4 \). Similarly, the complexity of \( H_i : \theta^1 > 2 \) is 1/2; that is, the complexity of the hypothesis \( H_i : \theta^* > 0 \), where \( \theta^* = \theta - 2 \). Stated otherwise, application of Theorem 2 to hypotheses of the form in Equation 4 implies that the parameter space is centered around \( d \) instead of 0.

As illustrated earlier, the complexity \( c_i \) of an informative hypothesis is independent of the prior distribution for the nuisance parameters. Furthermore, because the prior distribution for the nuisance parameter is specified to be vague, the posterior distribution and consequently the fit \( f_i \) of an informative hypothesis are virtually independent of the prior distribution for the nuisance parameters.

The overall conclusion is that the Bayes factor for the comparison of an informative hypothesis with its complement is independent of the prior distribution; that is, independent of subjective decisions that have to be made by a researcher, and thus, in this sense, objective. Moreover, the default prior specification in Mplus is suitable for computation of the Bayes factor in Equation 16.

**ESTIMATION OF \( c_i \) AND \( f_i \)**

According to Definition 2, complexity can be computed by the proportion of parameter space in agreement with the inequality-constrained hypothesis. For simple hypotheses, one can use \( Z = D! / B \). However, if more complex hypotheses are used, \( c_i \) can be obtained using sampling from the prior distribution \( h(\theta^1) \). In this situation, \( c_i \) can be obtained using two steps for \( t = 1, \ldots, T \):

1. Sample \( \theta^1_{tk} \) from \( \mathcal{N}(\theta^1_{tk} | \theta_0, \tau_0) \) for \( k = 1, \ldots, K \) and \( t = 1, \ldots, T \).
2. Estimate \( c_i \) by the proportion of \( \theta^1_t \) for \( t = 1, \ldots, T \) in agreement with \( H_i \).

The fit of the model, \( f_i \), can be estimated using a sample from the posterior distribution, \( g(\theta^1 | Y) \):

1. Sample \( \theta_{tk} \) from the posterior distribution for \( k = 1, \ldots, K \);
2. Estimate \( f_i \) by the proportion of \( \theta_t \) for \( t = 1, \ldots, T \) in agreement with \( H_i \).
Samples from $g(\theta^1|Y)$ can be obtained using Mplus using the option ANALYSIS: ESTIMATOR = BAYES in combination with the option SAVEDATA: BPARAMETERS ARE. These statements result in a file where all the parameters of the statistical model are saved for each iteration of the Gibbs sampler after omission of the burn-in phase. The second step of the computation of $f_i$ is to count the number of iterations where the inequality constraints are satisfied. One could do this by hand, but there is also an R package called `MplusAutomation` (Hallquist, 2003) that provides functions to compute these calculations based on Mplus output. For simple order restrictions $c_i$ can be computed by hand using Theorem 1, or it can be sampled using the procedure described earlier. Unfortunately, it is not yet possible to sample from the prior distribution in Mplus. The final step is to compute the $BF_{i\rightarrow}$ using Equation 16, and this can easily be done when the estimates for $c_i$ and $f_i$ are computed.

**PRECISION OF $c_i$ AND $f_i$**

In this section we evaluate the precision of the Bayes factor when evaluating inequality-constraint hypotheses. As shown in the previous section, three steps have to be executed to compute the required Bayes factor.

1. Obtain a sample of size $T$ from the unconstrained prior and posterior distribution.
2. Use this sample to determine $c_i$ and $f_i$ for the informative hypothesis under consideration; that is, determine the proportion of the sample from the unconstrained prior and posterior distribution in agreement with the hypotheses under consideration.
3. Compute the required Bayes factors.

This procedure will work fine if the number of parameters $K$ and the number of constraints $M$ are relatively small. For larger values of $K$ and $M$ the size of $T$ needed to obtain accurate estimates of $c_i$ and $f_i$ becomes so large that the procedure might no longer be of practical value. The question of how large $K$ and $M$ could be such that simply sampling from the prior and posterior of the unconstrained model will render accurate estimates of $c_i$ and $f_i$ is considered next.

The larger $M$ the smaller $c_i$ and the larger the sample that is needed to be able to accurately estimate $c_i$. Therefore the following elaboration is based on a complete ordering of the $K$ parameters; that is, $M = K - 1$ with $H \ell : \theta^1_1 > \ldots > \theta^1_K$. If $T$ is large enough to accurately estimate $c_i$ for this hypothesis, it will also be large enough to estimate $c_i$ for informative hypotheses based on fewer inequality constraints. In Table 1 the sizes $T$ of the sample from the prior distribution needed to be able to accurately estimate $c_i$ for $K = 2, \ldots, 8$ are displayed. Note that by accurately it is meant that the difference between the true $c_i$ and the lower bound and upper bound of a 95% central credibility interval for the estimate of $c_i$ is less than 10%. Assuming that $f_i$ is exactly known, this implies that $BF_{i\rightarrow} = f_i/c_i$ is never off by more than 10%.

The following can be concluded from Table 1. The complexity $c_i$ can accurately be evaluated for a complete ordering of six parameters using a sample of $T = 360,000$ from the prior distribution. The implication is that $c_i$ can almost exactly be computed for any inequality-constrained hypothesis based on $K = 6$ parameters using a sample of $T = 1,000,000$ from the prior distribution. Accurate computation of complexity for a complete ordering of seven parameters can be achieved using a sample of 5,000,000. For more than seven parameters,
accurate computation of $c_i$ for a complete ordering might not always be possible. Note that a complete ordering of parameters is not always of interest. Consider, for example, $0^1_1 > 0^1_2 > 0^1_3 > 0^1_4, 0^1_5 > 0^1_6, 0^1_7 > 0^1_8$, the combination of two independent sets of constraints. The complexity of this hypothesis is $1/24 \times 1/24 = .0017$. According to Table 1 a complexity of about .0017 can accurately be estimated using $T = 360,000$ and virtually exactly using $T = 1,000,000$. As exemplified earlier, Table 1 does not only apply to hypotheses in which the parameters at hand are completely ordered. As long as there is an idea of the size of the complexity to be estimated, Table 1 gives an indication of the sample sizes needed to be able to obtain an accurate estimate.

Now that the $T$ rendering accurate estimates of $c_i$ has been discussed, the implications for $f_i$ can be elaborated. If the sample from the prior is of size $T$, also use a sample from the posterior of size $T$. If the fit of a hypothesis is really bad, irrespective of the size of $T$, the number of parameter vectors sampled from the posterior distribution that are in agreement with $H_i$ will be very small. This is not a problem if one only wants to compare the hypothesis at hand to its complement. The conclusion will simply be that $H_i$ is not supported by the data.

REAL-LIFE EXAMPLE

To illustrate the Bayesian procedure of evaluating an inequality-constrained hypothesis versus its complement, we reevaluate the data of Boelen & van den Bout (2008; see also Boelen, Van de Schoot, van den Hout, van den Keijser, & van den Bout, 2010). The authors collected data to examine the distinctiveness of symptoms of CG and UG. Symptoms of CG represent distressing and disabling responses to bereavement, including persistent yearning, feeling that life is empty, and numbness. Symptoms of UG represent relatively benign grief reactions, including missing the lost person, crying, and having emotional recollections related to the death. To examine the distinctiveness of CG and UG, Boelen and van den Bout (2008) used data from 130 mourners who all received mental help after their loss. They were recruited via grief counselors, therapists, and other caretakers.

Symptoms of CG were assessed using the Dutch version of the Inventory of Complicated Grief (ICG; Boelen, van den Bout, de Keijser, & Hoijtink, 2003). This is a 30-item questionnaire tapping potentially problematic grief reactions. Respondents rate the presence of symptoms in the preceding month on 5-point scales ranging from never to all the time. Items representing UG were taken from the Present Feelings subscale of the Texas Revised Inventory of Grief (Faschingbauer, Zisook, & DeVaul, 1987). Respondents rate the presence of grief reactions...
on 5-point scales ranging from *completely true* to *completely false*. For the purpose of the illustrative analyses here we selected only three items from both questionnaires with the highest explained variance in the original factor analysis. Symptoms of depression were assessed using the Beck Depression Inventory (Beck, Steer, & Brown, 1996). Anxiety was assessed with the state version of the State–Trait Anxiety Inventory (Spielberger, 1983).

There is evidence that symptoms of CG are phenomenologically and qualitatively distinguishable from reactions representing UG with symptoms of CG but not UG being associated with concomitant mental health problems (Boelen et al., 2003; Dillen, Fontaine, & Verhofstadt-Denve, 2008). This is relevant, as it suggests that CG represents a clinical condition worthy of being included in psychiatric classification systems, such as the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000). Notably, CG is not yet included in these systems.

We reanalyzed these data, testing the prediction that, if CG symptoms are indeed more debilitating than symptoms of UG as prior research has shown, CG symptoms should be expected to be more strongly associated with symptoms of depression and anxiety than responses of UG. See Figure 1 for a graphical representation of the model where the indicators of the latent constructs are defined as categorical variables. For the *Mplus* syntax see Appendix C. The inequality-constrained hypothesis can be written as

$$A = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}, \quad \mu = \begin{bmatrix} \theta_1^l \\ \theta_2^l \\ \theta_3^l \\ \theta_4^l \end{bmatrix} \quad \text{and} \quad 0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (20)$$

![Figure 1](image)

**FIGURE 1** Path model between symptoms of complicated grief (CG), uncomplicated grief (UG), anxiety, and depression.
which renders:

$$H_i = \begin{bmatrix} \theta_1^1 - \theta_3^1 > 0 \\ \theta_2^1 - \mu_4 > 0 \end{bmatrix},$$

which is equivalent to $H_i : \theta_1^1 > \theta_3^1$ and $\theta_2^1 > \theta_4^1$.

Mplus v6.11 was used to estimate the model shown in Figure 1, using four chains, maximum likelihood estimates as starting values, and a Gelman–Rubin convergence criterion of .01 (see Appendix C). The default prior distributions of Mplus were used. The posterior results are shown in Table 2 (note that for reasons of space we did not include all thresholds) and the posterior distributions of $\theta_1^1 \ldots \theta_4^1$ are shown in Figure 2. For description purposes it can be interesting to inspect the credibility intervals. In Table 2 it can be seen that the 95% credibility interval of the four parameters of interest does contain zero. Moreover, it can be seen that the correlations between the independent and between the dependent variables are substantial. It is easy to show that $c_i = .25$ because $H_i$ consist of four equivalent hypotheses:

1. $H_1: \theta_1^1 > \theta_3^1$ and $\theta_2^1 > \theta_4^1$.
2. $H_2: \theta_1^1 < \theta_3^1$ and $\theta_2^1 > \theta_4^1$.
3. $H_3: \theta_1^1 > \theta_3^1$ and $\theta_2^1 < \theta_4^1$.
4. $H_4: \theta_1^1 < \theta_3^1$ and $\theta_2^1 < \theta_4^1$.

### TABLE 2

<table>
<thead>
<tr>
<th></th>
<th>Est.</th>
<th>SD</th>
<th>95% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG ($\theta_1^1$)</td>
<td>0.560</td>
<td>0.343</td>
<td>-0.099–1.280</td>
</tr>
<tr>
<td>UG ($\theta_2^1$)</td>
<td>0.083</td>
<td>0.356</td>
<td>-0.677–0.751</td>
</tr>
<tr>
<td>Depression only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG ($\theta_3^1$)</td>
<td>0.595</td>
<td>0.334</td>
<td>-0.046–1.310</td>
</tr>
<tr>
<td>UG ($\theta_4^1$)</td>
<td>0.086</td>
<td>0.354</td>
<td>-0.683–0.754</td>
</tr>
<tr>
<td>CG with UCG</td>
<td>0.821</td>
<td>0.069</td>
<td>0.654–0.919</td>
</tr>
<tr>
<td>Anxiety with Depression</td>
<td>0.700</td>
<td>0.081</td>
<td>0.514–0.804</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.472</td>
<td>0.237</td>
<td>3.011–3.943</td>
</tr>
<tr>
<td>Depression</td>
<td>3.535</td>
<td>0.242</td>
<td>3.071–4.020</td>
</tr>
<tr>
<td>Residual variance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>0.572</td>
<td>0.104</td>
<td>0.349–0.754</td>
</tr>
<tr>
<td>Depression</td>
<td>0.522</td>
<td>0.099</td>
<td>0.312–0.703</td>
</tr>
<tr>
<td>Factor loadings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.669</td>
<td>0.064</td>
<td>0.533–0.780</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.830</td>
<td>0.066</td>
<td>0.679–0.939</td>
</tr>
<tr>
<td>$\lambda_3$</td>
<td>0.598</td>
<td>0.081</td>
<td>0.421–0.739</td>
</tr>
<tr>
<td>$\lambda_4$</td>
<td>0.569</td>
<td>0.079</td>
<td>0.416–0.725</td>
</tr>
<tr>
<td>$\lambda_5$</td>
<td>0.550</td>
<td>0.095</td>
<td>0.345–0.716</td>
</tr>
<tr>
<td>$\lambda_6$</td>
<td>0.679</td>
<td>0.099</td>
<td>0.456–0.840</td>
</tr>
</tbody>
</table>

Note. Parameters of interest are shown in bold. Information about thresholds is omitted due to space constraints. CG = complicated grief; UG = uncomplicated grief.
FIGURE 2  Posterior distributions with posterior mean, mode, median, and 95% CI for $\theta_1^1$ \ldots $\theta_4^1$. (color figure available online)
Furthermore, for $f_i$ the proportion of iterations of the Gibbs sampler where the parameter estimates are in line with the inequality constraints can be computed using MplusAutomation. For our model, 82,800 iterations were performed and 57,934 of these iterations met the constraints, which renders an estimate for $f_i$ of .699. Using these results,

$$BF_{i \neg i} = \frac{f_i / c_i}{(1 - f_i)/(1 - c_i)} = \frac{.699/.25}{(1 - .699)/(1 - .25)} = 6.97. \quad (22)$$

In conclusion, $H_i$ is 6.97 times more likely than $H_{i \neg i}$. This finding indicates that the evidence for CG being a stronger predictor of depressive and anxious symptoms than UG is approximately seven times greater than the converse (i.e., UG is a stronger correlate of negative emotional symptoms).

CONCLUSION

Traditional hypotheses tests are not equipped to deal with informative hypotheses formulated in terms of inequality constraints among the parameters of a structural equation model. Also, the statistical tools that are available to evaluate an inequality-constrained hypothesis test it against the classical null hypothesis, against an unconstrained hypothesis, or against other inequality-constrained hypotheses, but not against its complement. Often, applied researchers are interested in the research questions “Is the hypothesis correct” and “Is the hypothesis incorrect?”, but tools to evaluate this hypothesis in SEM were not yet available. In this article, we presented a solution for this problem using Mplus.

Bayesian estimation was first made available in Mplus in 2010, which makes it possible for applied researchers to switch to Bayesian statistics easily. Also, the Bayesian toolkit in Mplus allows for the computation of a Bayes factor for the comparison of an informative hypothesis against its complement. The prior specifications are an important aspect of this comparison, but the defaults in Mplus are specified in such a way that the prior distributions are specified correctly. A next step would be to include the computation of the Bayes factor in the syntax Mplus so that researchers do not have to switch to other software to compute the Bayes factor. On the other hand, the computation is very simple and the R package MplusAutomation can compute the necessary estimates.

ACKNOWLEDGMENTS

Rens van de Schoot was supported by a grant from the Netherlands Organization for Scientific Research: NWO-VENI-451-11-008. Herbert Hoijtink was supported by a grant from the Netherlands organization for Scientific Research: NWO-VICI-453-05-002.

REFERENCES


**APPENDIX A**

**PROOF OF THEOREM 1**

If $A_1$ is divided in $D$ subsets of the same size, $\Theta^i = \{\Theta_1, \ldots, \Theta_D\}$, where each subset of $\Theta$ contains the same number of elements, there are $D!$ permutations of $\Theta: \Theta^i_1, \ldots, \Theta^{D!}$.

If the number of equivalent hypotheses is $Z$, $D! / Z = B$ of these permutations are in agreement with each equivalent hypothesis, that is, $[\Theta^1, \ldots, \Theta^{D!}] = [\ldots, \Theta^z_1, \ldots, \Theta^z_B, \ldots]$, leading to $H_{iz}: H_{iz1} \cup \ldots \cup H_{izB}$.

Because each $\Theta \in H_{iz}$ can be permuted to a $\Theta' \in H_{iz'}$ for which it holds that $h(\Theta) = h(\Theta')$ it holds that

$$c_{iz} = \int_{\Theta \in H_{iz}} h(\Theta) d\Theta = \int_{\Theta' \in H_{iz'}} h(\Theta') d\Theta' = c_{iz'}.$$  \hspace{1cm} (A.1)

Because $\sum_z c_{iz} = 1$ this implies that $c_{iz} = 1/Z$ for $z = 1, \ldots, Z$.

**APPENDIX B**

**PROOF OF THEOREM 2**

Let the rows of $A$ be denoted by $A_1, \ldots, A_M$, $A_m = [A_{m1}, \ldots, A_{mK}]$ and $d = [d_1, \ldots, d_M]$. Then

$$c_i = \frac{P(A_1\Theta - d_1 > 0, \ldots, A_M\Theta - d_M > 0)}{\int_{A_i\Theta - d_i > 0} … A_M\Theta - d_M > 0} \sim \mathcal{N}(V\Theta_0 - d, W^{-1} \tau_0^2),$$  \hspace{1cm} (B.1)
where

\[ V = \begin{bmatrix} \sum_k A_{1k} & \cdots & \sum_k A_{Mk} \\ \sum_k A_{1k} & \cdots & \sum_k A_{Mk} \\
\end{bmatrix}, \quad (B.2) \]

and

\[ W = \begin{bmatrix} \sum_k A_{1k}^2 & \cdots & \sum_k A_{Mk} A_{1k} \\ \sum_k A_{1k} A_{Mk} & \cdots & \sum_k A_{Mk}^2 \\
\end{bmatrix}. \quad (B.3) \]

Because \( \lim_{\xi \to 0} P(Z > 0 \mid Z \sim \mathcal{N}(V \theta_0 - d, W \tau_0^2)) = P(Z > 0 \mid Z \sim \mathcal{N}(0, W \tau_0^2)) = P(Z > 0 \mid Z \sim \mathcal{N}(0, W)) \), where \( Z \) is a vector of length \( M \) containing \( [A_1 \theta - d_1, \ldots, A_M \theta - d_M] \), the \( \lim_{\xi \to 0} c_i \) reduces to

\[ c_i = P(A_1 \theta - d_1 > 0, \ldots, A_M \theta - d_m > 0) \sim \mathcal{N}(0, W)). \quad (B.4) \]

that is, independent of \( \theta_0 \).

**APPENDIX C**

**MPLUS syntax**

DATA: FILE = data.dat;

VARIABLE:
NAMES ARE TRIG4 TRIG5 TRIG9 ITG4 ITG5 ITG10 STAITOT BDIM;
CATEGORICAL ARE TRIG4 TRIG5 TRIG9 ITG4 ITG5 ITG10;
MISSING ARE ALL (-999);

ANALYSIS:
ESTIMATOR = bayes; !here you select the Bayesian estimator
CHAINS = 4; !change the number of chains
PROCESSOR= 4 (starts); !increase computation speed
BCONVERGENCE = .01; !select a smaller convergence criterium
STVALUES = ml; !use maximum likelihood estimates as starting values

MODEL:
STAITOT BDIM on CG UCG;
UCG by TRIG4 TRIG5 TRIG9;
CG BY ITG4 ITG5 ITG10;

SAVEDATA: BPARAMETERS ARE c:/Bresults2.dat; !save the parameter estimates
for each iteration after burn-in