



A General Procedure for Testing Inequality Constrained Hypotheses in SEM

Leonard Vanbrabant,^{1,2} Rens Van de Schoot,^{2,3} Nancy Van Loey,⁴ and Yves Rosseel¹

¹Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University, Belgium

²Department of Methods & Statistics, Faculty of Social Sciences, Utrecht University, The Netherlands

³Optentia Research Program, North-West University, Vanderbijlpark, South Africa

⁴Association of Dutch Burn Centers, Department of Clinical and Health Psychology, Beverwijk, The Netherlands

Abstract: Researchers in the social and behavioral sciences often have clear expectations about the order and/or the sign of the parameters in their statistical model. For example, a researcher might expect that regression coefficient β_1 is larger than regression coefficients β_2 and β_3 . To test such a constrained hypothesis special methods have been developed. However, the existing methods for structural equation models (SEM) are complex, computationally demanding, and a software routine is lacking. Therefore, in this paper we describe a general procedure for testing order/inequality constrained hypotheses in SEM using the R package *lavaan*. We use the likelihood ratio (LR) statistic to test constrained hypotheses and the resulting plug-in p value is computed by either parametric or Bollen-Stine bootstrapping. Since the obtained plug-in p value can be biased, a double bootstrap approach is available. The procedure is illustrated by a real-life example about the psychosocial functioning in patients with facial burn wounds.

Keywords: (double) bootstrap, chi-square mixtures, informative hypothesis testing, LR statistic, order/inequality constraints

Structural equation modeling (SEM) software such as *lavaan* (Rosseel, 2012b) and *Mplus* (Muthén & Muthén, 2010) can be used to impose order/inequality constraints on the parameters of a statistical model. For example, there might be a hypothesis stating that regression coefficient β_1 is larger than regression coefficients β_2 and β_3 , which is denoted by

$$H : \beta_1 \geq \{\beta_2, \beta_3\}, \quad (1)$$

and is called an (order) constrained hypothesis (Barlow, Bartholomew, Bremner, & Brunk, 1972; Hoijtink, 2012; Klugkist, Laudy, & Hoijtink, 2005; Kuiper, Klugkist, & Hoijtink, 2010; Mulder, Hoijtink, & de Leeuw, 2012; Robertson, Wright, & Dykstra, 1988; Silvapulle & Sen, 2005; Van de Schoot, Hoijtink, Mulder, et al., 2011). In the literature, two methods are known for evaluating constrained hypotheses in SEM, namely the frequentist method proposed by Van de Schoot, Hoijtink, and Deković (2010) and the Bayesian method proposed by Van de Schoot, Hoijtink, Hallquist, and Boelen (2012). In this article, we focus on the frequentist procedure. However, the procedure is rather complex, since an abundant number of steps have to be carried out in

Mplus and R (R Development Core Team, 2015). Besides, the procedure is computationally demanding, limited to the parametric bootstrap, and no software routine is available.

Therefore, in the current paper we describe the R function *InformativeTesting()*. We will show that the *InformativeTesting()* function is easy to use and more flexible than the procedure described in Van de Schoot et al. (2010). Moreover, the *InformativeTesting()* function has some additional features, namely the Bollen-Stine bootstrap (Bollen & Stine, 1993) for non-normal data, parallel processing to reduce computational time, an option to produce high-quality plots based on the results, and the procedure does not depend on third-party commercial software but uses the open-source package *lavaan*.

The remainder of the paper is organized as follows. First, we describe the general structural equation model and its parameters on which constraints can be imposed. Furthermore, two hypothesis tests are introduced for testing constrained hypotheses and an illustration is presented to show how theoretical expectations can be converted into a constrained hypothesis. Second, we present a procedure

for testing constrained hypotheses. We introduce the parametric and Bollen-Stine bootstrap approaches and discuss the genuine double bootstrap method. Third, an overview of the `InformativeTesting()` function is presented. We show by means of a 5-step procedure how to convert the statistical model and the constraints into `lavaan` syntax and show how to set up the necessary function arguments. In addition, we describe the output of the `print()` and `plot()` methods using the results of the illustration. Finally, we make some concluding remarks.

Structural Equation Model With Constraints

A SEM with latent variables consists of two parts, namely a structural model and a measurement model. The structural model represents the structural equations that summarize the relationships between latent variables and can be written as:

$$\eta^g = \alpha^g + B^g \eta^g + \Gamma^g x^g + \zeta^g, \quad (2)$$

where the superscript g denotes group membership and runs from $g = 1, \dots, G$. The measurement model represents the link between the latent and observed variables and is written as:

$$y^g = v^g + \Lambda^g \eta^g + K^g x^g + \varepsilon^g, \quad (3)$$

where

$y \rightarrow p \times 1$ vector of dependent variables.

$\eta \rightarrow m \times 1$ vector of factors/latent variables.

v and $\alpha \rightarrow$ vectors of intercepts.

$\Lambda \rightarrow p \times m$ matrix of factor loadings.

K and $\Gamma \rightarrow$ matrices that contain slopes for exogenous covariates in the $(q \times 1)$ vector x .

$B \rightarrow m \times m$ vector of structural regression slopes.

ε and $\zeta \rightarrow$ vector of error terms.

$\Phi \rightarrow p \times p$ covariance matrix of ε .

$\Psi \rightarrow m \times m$ covariance matrix of ζ .

Furthermore, ε and ζ are multivariate normal distributed with means zero and covariance matrices Φ and Ψ , respectively.

The nonredundant free parameters of the model are collected in the parameter vector θ . Order/inequality constraints can be imposed on all¹ parameters of a structural equation model but in practice, only a subset of the free parameters are constrained. Then, let $\theta = \{\theta^a, \theta^b\}$,

where θ^a includes all parameters on which we impose constraints and where θ^b includes the remaining unconstrained parameters.

To test constrained hypotheses we consider two types of hypothesis tests, namely Type A and Type B (Silvapulle & Sen, 2005, pp. 61–62):

Type A:

$$\begin{aligned} H_{A0} : L\theta^a &= c \\ H_{A1} : L\theta^a &\geq c, \end{aligned} \quad (4)$$

Type B:

$$\begin{aligned} H_{B0} : L\theta^a &\geq c \\ H_{B1} : \theta^a &\in \mathbb{R}^k. \end{aligned} \quad (5)$$

If l is the number of inequality constraints imposed on θ^a , and k the number of parameters involved, then let L be an $l \times k$ matrix with known constants and c an $l \times 1$ vector with known constants (often this vector contains zeros). In hypothesis test Type A the null-hypothesis H_{A0} , in which all parameters are constrained to be equal, is tested against the constrained hypothesis H_{A1} . In hypothesis test Type B the constrained hypothesis H_{B0} is tested against the unconstrained model H_{B1} , which has no restrictions on θ^a . In order to find affirmative evidence for the constrained hypothesis, hypothesis test Type B plays a crucial role. Severe constraint violations result in rejecting the constrained hypothesis, since it is tested against the best fitting (i.e., unconstrained) hypothesis. Hypothesis test Type A is required to avoid false conclusions in case the inequality constraints are in fact equality constraints. In other words, hypothesis H_{A0} in test Type A should be rejected and the constrained hypothesis H_{B0} in test Type B not. If this is the case, loosely speaking this means that the constraints are in accordance with the data.

Illustration

To illustrate constrained hypothesis testing we use an example based on a cohort study in patients with facial burns (Hoogewerf, van Baar, Middelkoop, & van Loey, 2014). The example concerns a multiple group model with two groups, men and women. The sample consists of 77 respondents ($M_{\text{age}} = 39.95$, $SD = 14.05$) with facial burns, 78% of the respondents were men ($M_{\text{age}} = 38.96$, $SD = 13.76$) and 22% were women ($M_{\text{age}} = 44.04$, $SD = 14.81$). The aim of the study was to examine psychosocial functioning in patients with facial burn wounds. More in particular, in this part of the study the researchers wanted to test the hypothesis that the impact of burn severity on self-esteem would be higher in women

¹ In this article, we focus on imposing constraints on B , Γ , K , Λ , v , α , and Φ .

compared to men after controlling for symptoms of anxiety and depression. Burn severity was measured by the total body surface area burned (TBSA), which is the percentage of partial and full thickness burns on the total body. Anxiety and depression symptoms and self-esteem were measured using the Hospital Anxiety and Depression Scale (HADS; Spinhoven et al., 1997) and the Rosenberg's self-esteem scale (Rosenberg, 1965), respectively.

Previous studies have emphasized the greater importance of appearance on self-esteem and body image in women compared to men. One study (Strahan, Wilson, Cressman, & Buote, 2006) reported that women made more upward social comparisons than men on the body domain. In the aftermath of a burn injury that can cause lifelong disfigurement, it was empirically confirmed that female patients with burns are more dissatisfied with their appearance, leading to worse psychosocial functioning (Thombs et al., 2008). Women with facial burns in particular showed to be at higher risk for long-term depression symptoms (Wiechman et al., 2001). However, irrespectively of gender, depression symptoms are associated with low self-esteem and feelings of worthlessness (APA, 2000) and with maladaptive coping styles. Rumination (RUM) in particular has been strongly related to depression and anxiety symptoms (Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008). The theoretical assumptions between these variables are shown in Figure 1.

Since we are not interested in the intercepts, we can ignore the vectors α and \mathbf{v} in Equations 2 and 3. In the theoretical model we are dealing only with observed rather than latent variables. In the linear structural relationships (LISREL) tradition, all observed variables involved in a structural equation are upgraded to latent variables. Hence, the matrices Γ and K are not involved in estimating the model. Thus, we can write the model in Figure 1 as:

$$\begin{aligned}\eta^g &= B^g \eta^g + \zeta^g \\ \mathbf{y}^g &= \Lambda^g \eta^g + \varepsilon^g,\end{aligned}\quad (6)$$

where

$$B^g = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix}, \quad (7)$$

$$\Psi^g = \begin{bmatrix} \psi_{11} & & & & \\ 0 & \psi_{22} & & & \\ 0 & 0 & \psi_{33} & & \\ 0 & 0 & \psi_{43} & \psi_{44} & \\ 0 & 0 & \psi_{53} & \psi_{54} & \psi_{55} \end{bmatrix}, \quad (8)$$

and Λ^g is an identity matrix I and $\Phi^g = \mathbf{0}$. In our example $G = 2$, where $g = 1$ refers to men and $g = 2$ to women.

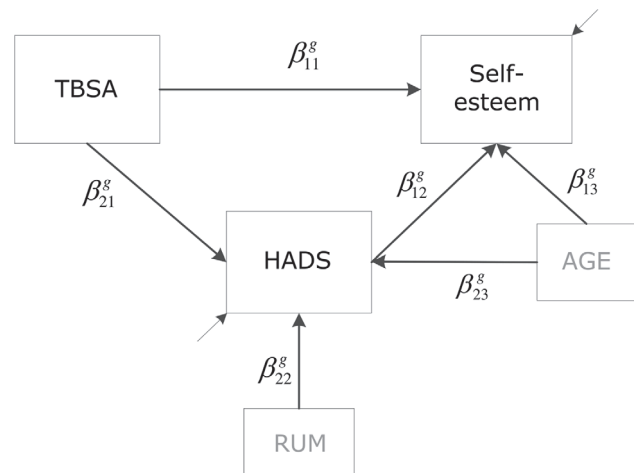


Figure 1. Multiple group SEM for the relation between total burned surface area (TBSA), self-esteem and symptoms of depression and anxiety (HADS) for men and women, controlling for age and coping style rumination.

Based on previous research on body image issues in patients with burns, the researchers hypothesized, first, that the impact of burn severity on self-esteem would be higher in women with facial burns compared to men with facial burns, after controlling for symptoms of depression and anxiety (HADS) and age. More precisely, the researchers expected a negative relation between TBSA and self-esteem for both men and women, and they expected the effect to be stronger for women. Those expectations can be converted directly into inequality and order constraints, namely $\beta_{11}^1 \leq 0$, $\beta_{11}^2 \leq 0$ and $\beta_{11}^2 \leq \beta_{11}^1$. Second, the researchers hypothesized a positive relation between TBSA and anxiety and depression symptoms for both men and women after controlling for rumination and age, and that the impact of TBSA on anxiety and depression symptoms would be higher in women. Therefore, the constraints for the second hypothesis are $\beta_{21}^1 \geq 0$, $\beta_{21}^2 \geq 0$ and $\beta_{21}^2 \geq \beta_{21}^1$. Using Equations 4 and 5, these constraints can be written as:

$$L = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad (9)$$

$$\theta^a = [\beta_{21}^1 \beta_{11}^1 \beta_{13}^1 \beta_{21}^2 \beta_{11}^2 \beta_{23}^2 \beta_{21}^2 \beta_{11}^2 \beta_{13}^2 \gamma_{21}^2 \beta_{22}^2 \beta_{23}^2]^T, \quad (10)$$

and $c = \mathbf{0}$. Then by multiplying matrix L by vector θ^a we can write the constrained hypothesis as:

$$H: \begin{bmatrix} -\beta_{11}^1 \geq 0 \\ -\beta_{11}^2 \geq 0 \\ \beta_{11}^1 - \beta_{11}^2 \geq 0 \\ \beta_{21}^1 \geq 0 \\ \beta_{21}^2 \geq 0 \\ -\beta_{21}^1 + \beta_{21}^2 \geq 0 \end{bmatrix} = \begin{bmatrix} \beta_{11}^1 \leq 0 \\ \beta_{11}^2 \leq 0 \\ \beta_{11}^2 \leq \beta_{11}^1 \\ \beta_{21}^1 \geq 0 \\ \beta_{21}^2 \geq 0 \\ \beta_{21}^2 \geq \beta_{21}^1 \end{bmatrix}. \quad (11)$$

In the next section we will discuss a procedure for testing such a constrained hypothesis. It is important to note that, the `InformativeTesting()` function does not require to construct the complex L matrix in Equation 9 manually. After the next section, we show that the constraints can be specified by a user-friendly text-based description.

Procedure for Testing Constrained Hypotheses in SEM

First, we start to discuss the parametric (Efron & Tibshirani, 1993, pp. 53–56) and the Bollen-Stine (Bollen & Stine, 1993, pp. 120–122) bootstrap approaches for obtaining a plug-in p value. Second, we introduce the genuine double bootstrap (Beran, 1988) method for adjusting the plug-in p value or α level.

Bootstrapping

An often-used procedure for comparing the fit of nested models, for example H_{A0} versus H_{A1} , is the likelihood ratio (LR) statistic for hypothesis test Type A. This is defined as:

$$LR = -2 \log \frac{L(\Sigma(\theta_{H_{A0}})Y)}{L(\Sigma(\theta_{H_{A1}})Y)}, \quad (12)$$

where L is the likelihood probability of the observed data Y as a function of $\Sigma(\theta)$ and where $\Sigma(\theta)$ is the estimated model implied covariance matrix under H_{A0} and H_{A1} . For hypothesis test Type B, H_{A0} and H_{A1} are replaced by H_{B0} and H_{B1} , respectively.

When the null and/or alternative hypothesis involves order/inequality constraints on the parameters, then the null-distribution of the LR statistic with multivariate normal data turns out to be the $\bar{\chi}^2$ -distributed (chi-square bar; Silvapulle & Sen, 2005). That is a weighted sum of

chi-squared distributions where the weights can be estimated via Monte Carlo simulations² or via the procedure described in Shapiro (1988) when dealing with linear regression and (only) linear constraints. Alternatively, the p value of the statistic can be computed directly via bootstrapping (Van de Schoot et al., 2010), which is called a plug-in p value and is denoted by \hat{p} .

This can be done by two types of bootstrap methods, namely by parametric (\hat{p}_{par}) and Bollen-Stine (\hat{p}_{bs}) bootstrapping.³ The plug-in p value usually refers to the parametric p value, however, for the sake of convenience we use the term plug-in p value also for the Bollen-Stine p value.

First, plug-in p value \hat{p}_{par} can be obtained by parametric bootstrapping and can be summarized by the following steps for a hypothesis test of Type A:

- Step 1. Estimate θ under the null-hypothesis H_{A0} using the observed data, resulting in $\Sigma(\theta_{H_{A0}})$. Also, estimate θ under H_{A1} resulting in $\Sigma(\theta_{H_{A1}})$ and compute the LR value for the observed data as shown in Equation 12, which is denoted by LR^{obs} .
- Step 2. Draw $B_r^1 = B_1^1, \dots, B_R^1$ bootstrap samples of size N from a known population distribution, say a multivariate normal distribution, using the estimated model implied covariance matrix $\Sigma(\theta_{H_{A0}})$. Superscript 1 denotes the first-level bootstrap samples.
- Step 3. Estimate θ_r for each bootstrap sample B_r^1 under H_{A0} and H_{A1} .
- Step 4. Compute for each B_r^1 sample the LR statistic. This results in a vector of R LR values, denoted by $LR_r^{boot} = LR_1^{boot}, \dots, LR_R^{boot}$.
- Step 5. To calculate the plug-in p value compute

$$\hat{p} = \frac{\sum_{r=1}^R I(LR_r^{boot} > LR^{obs})}{R}, \quad (13)$$

where I is the indicator function equaling 1 if the expression inside the brackets is true and 0 otherwise.

Parametric bootstrapping is a powerful method when the underlying assumption of the population distribution is satisfied. For example, for continuous data following a multivariate normal distribution. If this assumption holds the parametric bootstrap approach is expected to have a better accuracy (Gentle, Härdle, & Mori, 2004, p. 469). When this assumption is violated then the Bollen-Stine bootstrap approach would lead to more accurate results, since no underlying population distribution is assumed.

² Monte Carlo simulation is defined as a resampling technique that randomly generates samples from a known population distribution, such as the multivariate normal distribution (e.g., the parametric bootstrap). Nonparametric bootstrap procedures are similar to Monte Carlo simulations but the samples are drawn from the actual data and are therefore called resampling techniques (e.g., the Bollen-Stine bootstrap).

³ A third procedure exists, called the naive, or simple, bootstrap, but as shown in (Bollen & Stine, 1993, pp. 117–119) it is inaccurate for testing the LR statistic for structural equation models, since the bootstrap sample should only reflect sampling variability and possibly non-normality, but not model misfit.

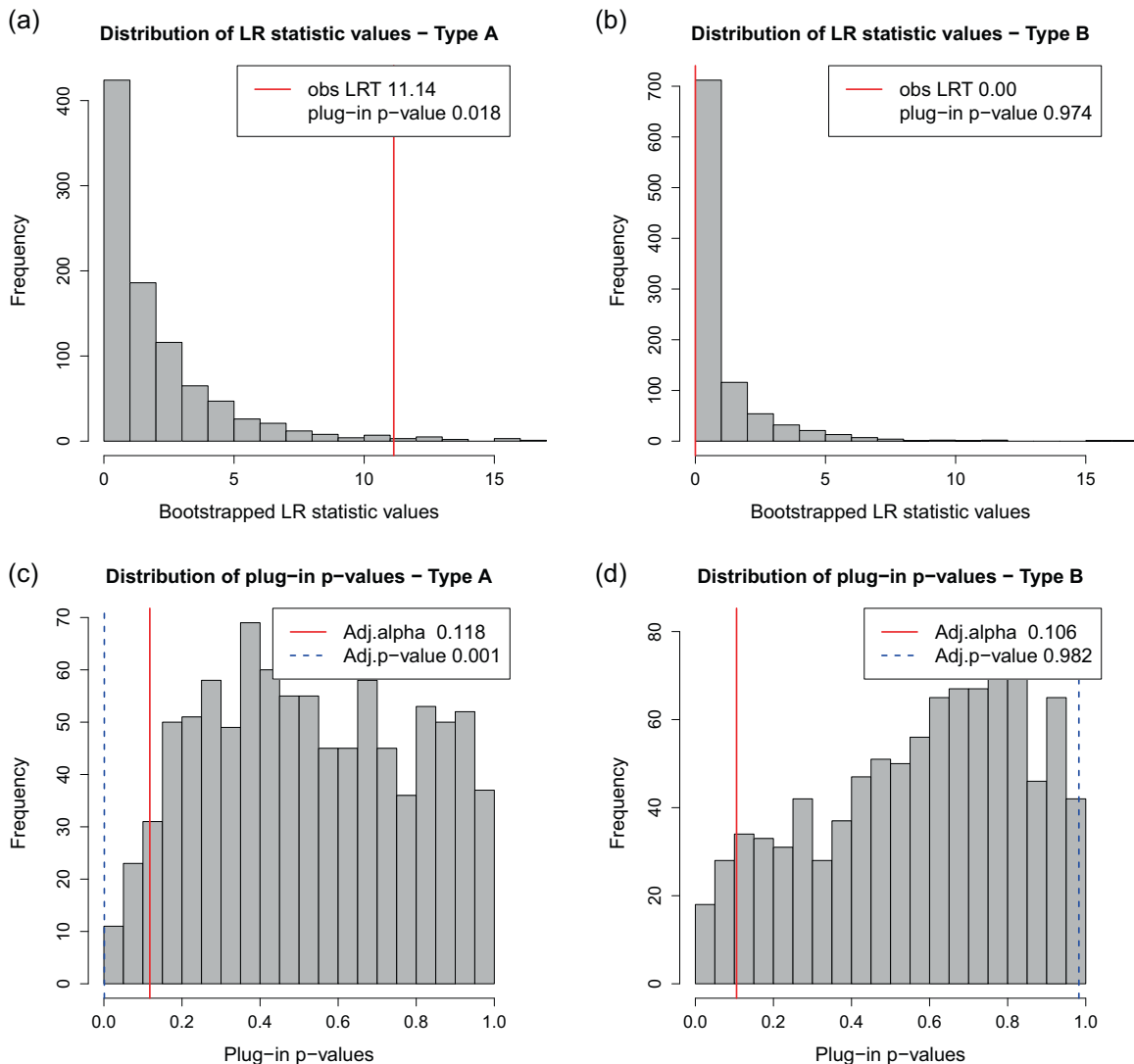


Figure 2. (a) Result of bootstrapping the LR values for the facial burns example for hypothesis test Type A. The solid line represents the LR^{obs} value. The proportion of LR^{boot} values on the right-hand side of the solid line is the plug-in p value \hat{p} . (b) See (a), but now for hypothesis test Type B. (c) Result of the genuine double bootstrap for the facial burns example for hypothesis test Type A. The non-uniform distribution of plug-in p values means that adjustment of α or \hat{p} is necessary. The solid line represents the adjusted alpha level α^* and the dashed line the adjusted plug-in p value \hat{p}^* . (d) See (c), but now for hypothesis test Type B.

The Bollen-Stine method is simply a nonparametric bootstrap where data are transformed in accordance with the null-hypothesis. Consequently, any non-normality of the data is preserved and therefore also retained in each bootstrap sample.

For computing the Bollen-Stine plug-in p value \hat{p}_{bs} only the first two steps are different compared to the parametric bootstrap:

- Step 1. Transform the observed data matrix so that its covariance structure is in accordance with the null-hypothesis.
- Step 2. Draw $B_r^1 = B_1^1, \dots, B_R^1$ bootstrap samples of size N from the transformed data and proceed with Step 3 from the parametric bootstrap approach.

To transform the data, we can use

$$Z = YS^{-1/2}\Sigma(\theta_{HA0})^{1/2}, \quad (14)$$

where Z is the transformed data, Y denotes the $N \times p$ data matrix of the centered observed variables, and S denotes the sample covariance matrix of Y (Bollen & Stine, 1993, p. 120).

Facial Burns Example Continued

For the facial burns example Figures 2a and 2b display the result of the bootstrap for hypothesis test Type A and Type B. Note that the result applies for both bootstrap

approaches. On the x -axis the LR values are given. Observe that most values are close to zero, since we sampled from the null-distribution. The solid vertical line represents the location of LR^{obs} . The plug-in p value is the proportion of LR^{boot} values on the right-hand side of the LR^{obs} . For a graphical representation of the parametric bootstrap, see Van de Schoot et al. (2010) and Van de Schoot and Strohmeier (2011). The two procedures previously described are repeated for hypothesis test Type B. However, estimating θ under the constrained hypothesis H_{B0} is more complex than under the equally constrained hypothesis H_{A0} . Computationally, linear equality constraints are generally easier to deal with than linear inequality constraints. Linear equality constraints result in a dimension reduction of the parameter vector. The resulting unconstrained problem can be solved using simpler methods for unconstrained optimization (Nocedal & Wright, 2006, ch. 17).

Double Bootstrapping

In the previous section we introduced the plug-in p value based on parametric and Bollen-Stine bootstrapping. A well-known property of the p value, and also of our plug-in p value, is that it is asymptotically a uniform distribution $[0, 1]$ under the H_0 , such that $P(p < \alpha | H_{A0}) = \alpha$. This is also true for \hat{p} when $R \rightarrow \infty$. However, when constraints are imposed on θ^a , it appears that $P(\hat{p} < \alpha | H_{A0}) \neq \alpha$. The parametric as well as the nonparametric bootstrap are not consistent if a parameter is on a boundary of the parameter space defined by (non)linear inequality constraints or a mixture between (non)linear inequality and equality constraints (Andrews, 2000). If this is the case either α needs to be adjusted or \hat{p} needs to be adjusted. Here, we discuss the genuine double bootstrap approach to adjust α and \hat{p} . We show how to obtain an adjusted α level, denoted by α^* and an adjusted plug-in p value, denoted by \hat{p}^* . In cases where it is necessary to make a distinction between the parametric and Bollen-Stine bootstrap we add the subscripts “par” and “bs” to \hat{p}^* or α^* .

For the genuine double bootstrap the following steps are needed for obtaining α^* or \hat{p}^* :

- Step 1. Draw $B_r^1 = B_1^1, \dots, B_R^1$ bootstrap samples of size N using either the parametric or Bollen-Stine bootstrap. Compute \hat{p}_{par} or \hat{p}_{bs} as defined in Equation 13.
- Step 2. Each of the B_r^1 is treated as an observed dataset from which second-level parametric or Bollen-Stine bootstrap samples $B_{rs}^2 = B_{r1}^2, \dots, B_{rs}^2$ are drawn.⁴

- Step 3. Use the second-level B_{rs}^2 bootstrap samples to compute R plug-in p values resulting in a vector of $\hat{p}_r = \hat{p}_1, \dots, \hat{p}_R$.

The result, so far, is a vector of R plug-in p values. The distribution of these plug-in p values should be uniform when $R \rightarrow \infty$. If this is the case then adjusting α or \hat{p} is not necessary. For the facial burns example the distributions for hypothesis tests Type A and Type B in Figures 2c and 2d are clearly not uniform and adjustment is necessary. For test Type A, $P(\hat{p} < .12 | H_{A0}) \neq .05$ and for test Type B, $P(\hat{p} < .11 | H_{B0}) \neq .05$. Now we continue with computing α^* or \hat{p}^* :

- Step 4a. The adjusted alpha level α^* is calculated by first ordering \hat{p}_r from small to large followed by computing the x th percentile, which is typically the 5th percentile at a significance level of 5%. In Figures 2c and 2d the solid lines show the adjusted α levels.

- Step 4b. According to Nankervis (2005), the adjusted p value \hat{p}^* is calculated by:

$$\hat{p}^* = \frac{\sum_{r=1}^R I(\hat{p}_r < \hat{p})}{R}. \quad (15)$$

In Figures 2c and 2d the dashed lines show the adjusted plug-in p values.

For a graphical representation of the genuine double bootstrap, see Van de Schoot et al. (2010). The steps previously described are repeated for hypothesis test Type B.

In the next section, we will discuss the `InformativeTesting()` function in more detail.

An Overview of the InformativeTesting Function in R Package lavaan

At the time of writing, the `InformativeTesting()` function is included in `lavaan`, a free and open-source R package for latent variable analysis (Rosseel, 2012a; <http://lavaan.org>).

Before we can test the constrained hypothesis of our facial burns example as defined in Equation 11 we need to go through five easy steps:

- Step 1 is to call the `lavaan` library:

```
R> library("lavaan")
```

⁴ The choice of S is always a tradeoff between precision and practical use. If we use as many as 1000 samples for both R and S , then we would need as many as 10^6 samples. Davison and Hinkley (2008, Ch. 5.6) suggest that $S = 249$ would be safe.

Step 2 is to load the observed data into R. The data can be a data frame containing the observed variables or a sample covariance full matrix with an optional mean vector. For our example the data are loaded into R as follows:

```
R> FacialBurns <- read.csv("burns.csv")
```

Step 3 is to convert the theoretical model in Figure 1 into lavaan syntax. The input model is specified by a text-based description called the lavaan model syntax and includes the overall model without constraints.

```
R> burnsModel <- 'Selfesteem ~ Age + HADS +
                  c(m1, f1)*TBSA
                  HADS ~ Age + RUM +
                  c(m2, f2)*TBSA '
```

where $m1$, $f1$, $m2$, and $f2$ are arbitrary labels which are necessary for imposing the constraints.

Step 4 is to convert the constraints into lavaan syntax. For the sake of convenience we do not use the greek letter β with subscripts and superscript, but simple labels. Therefore let $\beta_{11}^1 = m1$, $\beta_{11}^2 = f1$, $\beta_{21}^1 = m2$, and $\beta_{21}^2 = f2$. The constraints are specified by a text-based description, called the lavaan constraints syntax and describe the linear order/inequality constraints imposed on the model. A major advantage of this text-based description is that users do not have to specify the complex L matrix (see 9) themselves. Then, the constraints are defined as follows:

```
R> burnsConstraints <- 'm1 < 0
                       f1 < 0
                       f1 < m1
                       m2 > 0
                       f2 > 0
                       f2 > m2 '
```

Note that these constraints are equal to the right-hand side of Equation 11. Also note that, it is only necessary to specify the overall model and the constraints, the equality constrained model H_{A0} and the unconstrained model H_{B1} are generated automatically by the InformativeTesting() function and thus need not to be specified. For more information about how to create the model and constraints syntax, see the lavaan manual (Rosseel, 2012a).

Step 5 is to set up the necessary InformativeTesting() function arguments. For an overview of all function arguments, see ?InformativeTesting. The first argument to InformativeTesting() is the model defined in Step 3. The second argument is the

observed data. The third argument is the constraints imposed on the model in Step 4. The fourth and fifth arguments define the number of bootstrap draws and the number of double bootstrap draws, respectively. In our example, we used $R = 1,000$ and $S = 249$. The group argument specifies the grouping variable, which in our case the variable name "Sex" in the data frame. The parallel and ncpus arguments are needed to use parallel processing. In this example, we used eight cores for the computations.

```
R> burnsIT <- InformativeTesting(model =
                                burnsModel, data = FacialBurns,
                                constraints = burnsConstraints,
                                double.bootstrap.R = 249,
                                R = 1000, group = "Sex",
                                parallel = "multicore", ncpus = 8)
```

The InformativeTesting() function bootstraps LR values and it returns an object of class "InformativeTesting" for which a plot() method is available, which is discussed later. By default the InformativeTesting() function uses the Bollen-Stine bootstrap approach (type = "bollen.stine") and the genuine double bootstrap for adjusting the plug-in p value (double.bootstrap = "standard"). However, users can easily switch to the parametric bootstrap (type = "parametric") or turn the double bootstrap off (double.bootstrap = "no"). Furthermore, by default the InformativeTesting() function generates $R = 1,000$ bootstrap draws and returns a vector with the bootstrapped LR values (return.LRT = TRUE). For the genuine double bootstrap double.bootstrap.R = 249 double bootstrap samples are drawn. Note that for the "standard" double bootstrap by default $S = 249$ and a significance level of 5% is used to compute α^* (double.bootstrap.alpha = 0.05).

In the next section we will discuss the print() and plot() methods for the InformativeTesting() function using the facial burns example.

Facial Burns Example Continued: print() and plot()

Perhaps the most informative method to view the results is plot(). The plot() function plots the distributions of the bootstrapped LR values and also the distributions of the plug-in p values in case of the genuine double bootstrap. The plot() method can be called without additional arguments to plot all available plots.

Separate plots for the distribution of LR values or the plug-in p values can be requested. For the distribution of LR values the argument `type = "lr"` is added (see Figures 2a and 2b) and for the distribution of plug-in p values the argument `type = "ppv"` is added, see Figures 2c and 2d. The first argument to `plot()` is the returned object from the `InformativeTesting()` function.

```
R> plot(burnsIT)
```

For the `plot()` results for the facial burn example see Figures 2a–2d. The default plot arguments can be overruled by the user to adjust the plots. For example, the axes labels, main title, number of breaks, and colors can be adjusted. For all available options, see `?plot.InformativeTesting`.

A table of the `print()` function is displayed with the results of the facial burn example.

```
R> burnsIT
```

```
InformativeTesting:      Order/Inequality
Constrained Hypothesis Testing:
```

```
Variable names in model : Self-esteem HADS
                        Age TBSA RUM
Number of variables     : 5
Number of groups        : 2
Used sample size per group: 60 17
Used sample size        : 77
Total sample size       : 118
Estimator               : ML
Missing data            : listwise
Bootstrap method        : bollen.stine
Double bootstrap method : standard
```

Type A test:

```
H0: all restrictions active (=) vs.
H1: at least one restriction strictly true (>)
```

```
Test statistic: 11.1374,
adjusted p-value: 0.0011 (alpha = 0.05)
unadjusted p-value: 0.0182 (alpha = 0.1176)
```

Type B test:

```
H0: all restrictions true vs.
H1: at least one restriction false
```

```
Test statistic: 0.0000,
adjusted p-value: 0.9824 (alpha = 0.05)
unadjusted p-value: 0.9742 (alpha = 0.1055)
```

The results for hypothesis test Type A, see also Figures 2a and 2c, show that the equality constrained hypothesis H_{A0} is rejected ($LR^{obs} = 11.1374$, $\hat{p}^* = .001$, $\alpha = .05$). In other words, the observed LR statistic $LR^{obs} = 11.1374$, see solid line

Figure 2a, is more extreme than we would expect by chance. The adjusted plug-in p value \hat{p}^* is the proportion of plug-in p values on the right-hand side of the dashed line. The results for Type B, see also Figures 2b and 2d, show that the constrained hypothesis H_{B0} cannot be rejected ($LR^{obs} = 0$, $\hat{p}^* = .982$, $\alpha = .05$). The observed LR statistic, $LR^{obs} = 0$, indicates that the constrained hypothesis does not fit significantly worse than the unconstrained. More precisely, an LR value of 0 indicates that no constraints are violated. Therefore, we can conclude that the results show strong evidence for the constrained hypothesis stated in Equation 11. In other words, there exists a negative relation between TBSA and self-esteem for both men and women, and the relation is stronger for women. It is also confirmed that the relation between TBSA and HADS is positive for both men and women and that the relation is stronger for women. A visual inspection of the constrained or unconstrained model parameters reinforces our conclusion. This can be done as follows for the unconstrained model. Some parts of the output are removed due to its length.

```
R> summary(burnsIT$fit.B1)
```

```
Group 1 [1]:
Regressions:
                Estimate
Selfesteem ~
Age              0.019
TBSA      (m1)  -0.148
HADS            -0.475
HADS ~
Age              0.070
TBSA      (m2)   0.143
RUM             1.036
```

```
Group 2 [2]:
```

```
Regressions:
                Estimate
Selfesteem ~
Age              0.106
TBSA      (f1)  -0.241
HADS            -0.453
HADS ~
Age             -0.010
TBSA      (f2)   0.254
RUM             -0.656
```

The constrained parameter estimates can be requested as follows:

```
R> summary(burnsIT$fit.A1)
```

A tutorial function with the R-code and the data from the facial burns example is available online at <https://github.com/LeonardV/InformativeTesting>. For more information and a gentle introduction to informative hypotheses,

we recommend the book of Hoijtink (2012), and also the papers from Van de Schoot, Hoijtink, and Romeijn (2011), Van de Schoot and Strohmeier (2011), and Van de Schoot and Wong (2011) for more applied examples.

Concluding Remarks

In classical null-hypothesis testing researchers can only find indirect evidence for their specific hypotheses if the null-hypothesis is rejected. Therefore, we believe that evaluating informative hypotheses, by means of imposing order/inequality constraints on the parameters of a statistical model, allow researchers to directly evaluate their expectations and get more insightful results compared to testing the classical null-hypothesis against catch-all rivals. In addition, Vanbrabant, Van de Schoot, and Rosseel (2015; and the references therein) have shown that substantial power can be gained when an increasing number of order/inequality constraints are included into the hypothesis. Researchers who are dealing with small samples in particular may benefit from this power gain.

Hypothesis test Type A can reject the null-hypothesis H_{A0} even if the alternative hypothesis H_{A1} is violated by the data. Rejecting H_{A0} does not mean that H_{A1} is true. Note that, this applies also to classical null-hypothesis testing. The power of hypothesis tests Type A and Type B is centered in the alternative hypothesis H_1 . It is only under H_0 that their type I errors are close to the nominal level. In spite of this, hypothesis test Type A can be useful since hypothesis test Type B cannot make a distinction between equality and inequality constraints. Hypothesis test Type B plays a crucial role in constraint misspecification.

The `InformativeTesting()` function discussed in this paper is the first software routine for testing order/inequality constrained hypotheses in SEM. If researchers want to test a constrained hypothesis, then the procedure with the `InformativeTesting()` function for `lavaan` is easier to use and faster compared to the procedure proposed in Van de Schoot et al. (2010).

However, testing constrained hypotheses in SEM is due to the genuine double bootstrap procedure computationally very expensive. Therefore, computational time remains a limitation and procedures to decrease it further are investigated.

Furthermore, we conducted a small simulation study in which we investigated the performance of the Bollen-Stine and parametric bootstrap approaches in terms of type I errors ($\alpha = .05$). We have set up a simulation design in which we varied the sample size and the normality of the data. We chose a small sample size of $N = 50$ and a large sample size of $N = 500$. We generated normal and very non-normal data with a skewness of 1.50 and a kurtosis

of 3.75. The results show that the Bollen-Stine bootstrap outperforms the parametric bootstrap in case of non-normal data. We recommend to use the parametric bootstrap only in case of normal distributed samples. Currently, the parametric bootstrap for the `InformativeTesting()` function is only valid for continuous data following a multivariate normal distribution.

Finally, we advise to use the genuine double bootstrap and only to switch the double bootstrap off when exploration is the goal of the analysis.

Acknowledgments

The first author is a PhD fellow of the Research Foundation Flanders (FWO) at Ghent university (Belgium) and at Utrecht University (The Netherlands). The second author is supported by a grant from the Netherlands Organization for Scientific Research: NWO-VENI-451-11-008. Data for the empirical sample were obtained for a study funded by the Dutch Burns Foundation (Grant No. 05.109). We thank all participating patients and the research team in the respective burn centers (Groningen: M. Bremer, G. Bakker; Beverwijk: A. Boekelaar; Rotterdam: A. van de Steenoven, H. Hofland).

References

- Andrews, D. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica*, 68, 399–405.
- APA. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Association.
- Barlow, R. E., Bartholomew, D. J., Bremner, H. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions*. New York, NY: Wiley.
- Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83, 687–697. doi: 10.1080/01621459.1988.10478649
- Bollen, K. A., & Stine, R. (1993). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods and Abstracts*, 21, 205–229. doi: 10.1177/0049124192021002004
- Davison, A., & Hinkley, D. (2008). *Bootstrap methods and their application* (10th ed.). Cambridge, UK: Cambridge University Press.
- Efron, E., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.
- Gentle, J., Härdle, W., & Mori, Y. (Eds.). (2004). *Handbook of computational statistics: Concepts and methods*. Berlin, Germany: Springer.
- Hoijtink, H. (2012). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: Taylor & Francis.
- Hoogewerf, C. J., van Baar, M. E., Middelkoop, E., & van Loey, N. E. (2014). Impact of facial burns: Relationship between depressive symptoms, self-esteem and scar severity. *General Hospital Psychiatry*, 36, 271–276.
- Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach.

- Psychological Methods*, 10, 477–493. doi: 10.1037/1082-989X.10.4.477
- Kuiper, R., Klugkist, I., & Hoijtink, H. (2010). A Fortran 90 program for confirmatory analysis of variance. *Journal of Statistical Software*, 34, 1–31.
- Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, 46, 1–38.
- Muthén, L., & Muthén, B. (2010). *Mplus: Statistical analysis with latent variables: User's guide*. Los Angeles, CA: Muthén & Muthén.
- Nankervis, J. (2005). *Stopping rules for double bootstrap tests*. (Working paper, University of Essex).
- Nocedal, J., & Wright, S. (2006). In V. Mikosch, S. Resnick, & S. Robinson (Eds.), *Numerical optimization* (2nd ed.). New York, NY: Springer.
- Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3, 400–424. doi: 10.1111/j.1745-6924.2008.00088.x
- R Development Core Team. (2015). *R: A language and environment for statistical computing* [Computer software manual] (ISBN 3-900051-07-0). Vienna, Austria: R Core Team.
- Robertson, T., Wright, F. T., & Dykstra, R. L. (1988). *Order restricted statistical inference*. New York, NY: Wiley.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosseel, Y. (2012a). *lavaan: An R package for structural equation modeling and more* [Computer software manual]. Ghent, Belgium: Department of Data Analysis Ghent University. Retrieved from <http://lavaan.org>
- Rosseel, Y. (2012b). *lavaan: An R package for structural equation modeling (version 0.4–13)*. *Journal of Statistical Software*, 48, 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, 56, 49–62. doi: 10.2307/1403361
- Silvapulle, M., & Sen, P. (2005). *Constrained statistical inference: Order, inequality, and shape constraints*. Hoboken, NJ: Wiley.
- Spinhoven, P., Ormel, J., Sloekers, P. P. A., Kempen, G., Speckens, A. E. M., & VanHemert, A. M. (1997). A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychological Medicine*, 27, 363–370. doi: 10.1017/S0033291796004382
- Strahan, E. J., Wilson, A. E., Cressman, K. E., & Buote, V. M. (2006). Comparing to perfection: How cultural norms for appearance affect social comparisons and self-image. *Body Image*, 3, 211–227. doi: 10.1016/j.bodyim.2006.07.004
- Thombs, B. D., Notes, L. D., Lawrence, J. W., Magyar-Russell, G., Bresnick, M. G., & Fauerbach, J. A. (2008). From survival to socialization: A longitudinal study of body image in survivors of severe burn injury. *Journal of Psychosomatic Research*, 64, 205–212. doi: 10.1016/j.jpsychores.2007.09.003
- Van de Schoot, R., Hoijtink, H., & Deković, M. (2010). Testing inequality constrained hypotheses in SEM models. *Structural Equation Modeling*, 17, 443–463. doi: 10.1080/10705511.2010.489010
- Van de Schoot, R., Hoijtink, H., Hallquist, M., & Boelen, P. (2012). Bayesian evaluation of inequality-constrained hypotheses in SEM models using mplus. *Structural Equation Modeling*, 19, 593–609. doi: 10.1080/10705511.2012.713267
- Van de Schoot, R., Hoijtink, H., Mulder, J., Van Aken, M. A. G., Orobio de Castro, B., Meeus, W., & Romeijn, J. (2011). Evaluating expectations about negative emotional states of aggressive boys using Bayesian model selection. *Developmental Psychology*, 47, 203–212. doi: 10.1037/a0020957
- Van de Schoot, R., Hoijtink, H., & Romeijn, J. (2011). Moving beyond traditional null hypothesis testing: Evaluating expectations directly. *Frontiers in Psychology*, 2, 1–5. doi: 10.3389/fpsyg.2011.00024
- Van de Schoot, R., & Strohmeier, D. (2011). Testing informative hypotheses in SEM increases power: An illustration contrasting classical hypothesis testing with a parametric bootstrap approach. *International Journal of Behavioral Development*, 35, 180–190. doi: 10.1177/0165025410397432
- Van de Schoot, R., & Wong, T. (2011). Do antisocial young adults have a high or a low level of self-concept? *Self and Identity*, 11, 148–169. doi: 10.1080/15298868.2010.517713
- Vanbrabant, L., Van de Schoot, R., & Rosseel, Y. (2015). Constrained statistical inference: Sample-size tables for anova and regression. *Frontiers in Psychology*, 5, 1–8. doi: 10.3389/fpsyg.2014.01565
- Wiechman, S. A., Ptacek, J., Patterson, D. R., Gibran, N. S., Engrav, L. E., & Heimbach, D. M. (2001). Rates, trends, and severity of depression after burn injuries. *The Journal of Burn Care & Rehabilitation*, 22, 417–424. doi: 10.1097/00004630-200111000-00012

Received November 12, 2014
 Revision received August 26, 2016
 Accepted September 16, 2016
 Published online June 2, 2017

Leonard Vanbrabant is a PhD student at the Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University. His research interests focus on constrained statistical inference and structural equation modeling, and applications related to post-traumatic stress symptoms and burns.

Rens van de Schoot obtained his PhD from Utrecht University, The Netherlands. He is an associate professor at the Department of Methods and Statistics, Utrecht University. His research interests include Informative hypothesis testing, Bayesian statistics, and structural equation modeling, and applications related to post-traumatic stress disorder.

Nancy Van Loey is Senior Researcher at the Association of Dutch Burn Centres, Beverwijk, The Netherlands. She is involved in research focused on psychological problems and quality of life following burn injuries.

Yves Rosseel obtained his PhD from Ghent University, Belgium. He is now an associate professor at the Department of Data Analysis, Faculty of Psychology and Educational Sciences, Ghent University. His research interests include computational statistics, connectivity and causality, and structural equation modeling. He is the author of lavaan, an R package for structural equation modeling.

Leonard Vanbrabant

Department of Data Analysis
 Faculty of Psychology and Educational Sciences
 Ghent University
 Henri Dunantlaan 1
 9000 Ghent
 Belgium
leonard.vanbrabant@ugent.be