Article

BIC Extensions for Order-constrained Model Selection

J. Mulder[1,2] and A. E. Raftery[3]

Abstract

The Schwarz or Bayesian information criterion (BIC) is one of the most widely used tools for model comparison in social science research. The BIC, however, is not suitable for evaluating models with order constraints on the parameters of interest. This article explores two extensions of the BIC for evaluating order-constrained models, one where a truncated unit information prior is used under the order-constrained model and the other where a truncated local unit information prior is used. The first prior is centered on the maximum likelihood estimate, and the latter prior is centered on a null value. Several analyses show that the order-constrained BIC based on the local unit information prior works better as an Occam's razor for evaluating order-constrained models and results in lower error probabilities. The methodology based on the local unit information prior is implemented in the R package "BICpack" which allows researchers to easily apply the method for order-constrained model selection. The usefulness of the methodology is illustrated using data from the European Values Study.

Keywords

Bayesian information criterion, order constraints, truncated priors, European Values Study, model selection

[1] Tilburg University, The Netherlands
[2] Jheronimus Academy of Data Science, Hertogenbosch, The Netherlands
[3] University of Washington, Seattle, WA, USA

Corresponding Author:
J. Mulder, Tilburg University, Warandelaan 1, Noord Brabant, 5000 LE Tilburg, the Netherlands.
Email: j.mulder3@tilburguniversity.edu

Introduction

The Bayesian information criterion (BIC) is one of the most commonly used model evaluation criteria in social research, for example, for categorical data (Raftery 1986), event history analysis (Vermunt 1997), or structural equation modeling (Lee and Song 2007; Raftery 1993). The BIC, originally proposed by Schwarz (1978), can be viewed as a large sample approximation of the marginal likelihood (Jeffreys 1961) based on the so-called unit information prior. This unit information prior contains the same amount of information as would a typical single observation (Raftery 1995).

The BIC has several useful properties. First, it can be used as a default quantification of the relative evidence in the data between two statistical models. Second, it can straightforwardly be used for evaluating multiple statistical models simultaneously. Third, it is consistent for most well-behaved problems in the sense that the evidence for the true model converges to infinity (Kass and Wasserman 1995). Fourth, it behaves as an Occam's razor by balancing model fit (quantified by the log likelihood function at the maximum likelihood estimate [MLE]) and model complexity (quantified by the number of free parameters). Fifth, it is easy to compute using standard statistical software: Only the MLEs, the maximized loglikelihood, the sample size, and the number of model parameters are needed to compute it. All these useful properties have contributed to the popularity and usefulness of the BIC in social research.

Despite the general applicability of the BIC, it is not suitable for evaluating statistical models with order constraints on certain parameters. In a regression model, for instance, it may be expected that the first predictor has a larger effect on the outcome variable than the second predictor, and the second predictor is expected to have a larger effect than the third predictor. This can be translated to the following order-constrained model, $M_1 : \beta_1 > \beta_2 > \beta_3$, where β_k denotes the effect of the kth predictor on the outcome variable. This model can then be tested against conflicting models, such as a model with competing order constraints, for example, $M_2 : \beta_3 > \beta_1 > \beta_2$, a model where the effects are expected to be equal, $M_3 : \beta_1 = \beta_2 = \beta_3$, or the complement of these models, denoted by M_4. Under the complement model M_4, the true values for the β s do not satisfy any of the constraints under models M_1, M_2, or M_3.

The reason that the BIC is not suitable for testing models with order constraints is that the number of free parameters does not properly capture the complexity of a model. In the above model M_1, all three β parameters are free parameters, but saying that M_1 is equally complex as a model with no

constraints, that is, $(\beta_1, \beta_2, \beta_3) \in \mathbb{R}^3$, seems incorrect. Furthermore, the BIC is based on the Laplace approximation of the marginal likelihood. It is as yet unclear how well the approximation performs in the case of models with order constraints. The complicating factor is that the approximation assumes that the maximum value of the integrand is an interior point of the integrated region. This assumption is violated if the maximum likelihood (or posterior mode) does not lie in the integrated region.

Testing order constraints is particularly useful because effect sizes can only be interpreted relative to each other in the study and relative to the field of research (J. Cohen 1988). An effect size of, say, .3, of educational level on attitude toward immigrants might seem substantial for a sociologist, while .3 might not be interesting when it quantifies the effect of a medical treatment on the amount of pain of a patient. Thus, instead of interpreting the magnitudes of effects by their estimated values, it may be more informative to interpret them relative to each other, as is done using order-constrained model selection. This would allow us to assess which effects dominate other effects in the study. Further, order-constrained model selection is useful when testing scientific expectations that can be formulated using order constraints. Examples will be presented in the second section in sociological applications, but there is also an increasing body of literature on this topic (e.g., Böing-Messing et al. 2017; Braeken, Mulder, and Wood 2015; de Jong, Rigotti, and Mulder 2017; Hoijtink 2011; Klugkist, Laudy, and Hoijtink 2005; Kluytmans et al. 2012; Mulder and Fox 2019; Mulder and Pericchi 2018; van de Schoot et al. 2012). By testing order-constrained models, we can quantify the evidence in the data for one scientific theory against others. Order-constrained models are also naturally specified when one is interested in the effect of an ordinal categorical variable on an outcome variable of interest.

It is also important to note that the inclusion of order constraints results in more statistical power. This can be explained by the smaller subspace for the parameters under an order-constrained model compared to an equivalent model without the order constraints. The order constraints make the model "less complex," resulting in a smaller penalty for model complexity, and thus in more evidence for an order-constrained model that is supported by the data.

In this article, we explore how the BIC can be extended to enable order-constrained model selection. First, a unit information prior is considered that is truncated in the order-constrained subspace. This results in a BIC that may not properly incorporate the relative complexity of an order-constrained

model. For this reason, an alternative local unit information prior is considered which is centered on a null value. This prior results in a BIC that properly incorporates the relative fit and complexity of order-constrained models. The R package "`BICpack`" has been developed for order-constrained model selection in popular models such as generalized linear models, survival models, and ordinal regression models.

To our knowledge, there have been two other proposals for the BIC for evaluating models with order (or inequality) constraints by Romeijn, van de Schoot, and Hoijtink (2012) and Morey and Wagenmakers (2014), and we will compare our proposal to theirs. The article is organized as follows. We motivate the evaluation of statistical models with order constraints on the parameters of interest in the context of the European Values Study in the second section. In the third section, we discuss BIC approximations of the marginal likelihood under an order-constrained model. Fourth section provides a numerical evaluation of the methods, while fifth section describes software to implement the methods. Sixth section explains how to apply the new method for testing social theories in the European Values Study, and seventh section discusses the results.

Order-constrained Model Selection in Social Research

In this section, we present two situations where order-constrained model selection is useful. First, theories often make an assumption about the relative importance of certain predictors on an outcome variable. This can be formalized by specifying order constraints on the effects of these predictor variables. We will show this in Application 1 using Ethnic Competition Theory (Scheepers, Gijsberts, and Coenders 2002). Second, a researcher may have an expectation about the direction of an effect of a predictor variable with an ordinal measurement level. When modeling this ordinal predictor variable using dummy variables, the expected directional effect can be translated to a set of order constraints on the effects of these dummy variables. This will be shown in Application 2 by considering Inglehart's Generational Replacement Theory.

Application 1: Assessing the Importance of Different Dimensions of Socioeconomic Status

In most European countries, the majority of immigrants are located in the lower strata of society. For this reason, lower-strata members of the European majority population who hold similar social positions as the ethnic minorities, having a relatively low social class, low educational level, or

low-income level will on average compete more with ethnic minorities than will other citizens in the labour market. Therefore, Ethnic Competition Theory (Scheepers et al. 2002) would predict that higher social class, educational level, or income level would result in a more positive attitude toward immigrants. Furthermore, it is likely that social class (which reflects the type of job a person has) has the largest impact because one's social class is directly related to the labour market. The effects of education are less direct, and therefore, it is expected that one's educational level has a lower impact on attitude toward immigrants than social class. Finally, it would be expected that the effect of income would be the lowest but still positive. This expectation will be formalized in model M_1 which is provided below.

Alternatively, due to the importance of education in shaping one's identity (A. K. Cohen et al. 2013; van der Waal, de Koster, and ten Kate 2015), it might be expected that education is the most important factor explaining one's attitude toward immigrants, followed by social class and income for which no specific ordering is expected (formalized in M_2). A third hypothesis is that all three dimensions have an equal and positive effect on attitudes toward immigrants (M_3). Finally, it may be that none of these three hypotheses is true (M_4).

To evaluate these expectations, we first write down the linear regression model where the attitude toward immigrants is the outcome variable, and social class, educational level, and income are the predictor variables while controlling for age. The ith observation is modeled as follows:

$$\begin{aligned}
\text{attitude}(i) = \theta_0 &+ \text{class}(i) \times \theta_{\text{class}} + \text{education}(i) \times \theta_{\text{education}} \\
&+ \text{income}(i) \times \theta_{\text{income}} + \text{gender}(i) \times \theta_{\text{gender}} + \text{error}(i),
\end{aligned} \quad (1)$$

for $i = 1, \ldots, n$. The predictor variables are all standardized. In equation (1), θ_{class}, $\theta_{\text{education}}$, and θ_{income} are the standardized coefficients for social class, educational level, and income, respectively, θ_{gender} is the standardized coefficient for gender, and the errors are assumed to be independent and normally distributed with unknown variance.

The four expectations given above can be formalized using competing statistical models with different order constraints on the standardized effects, namely,

$$\begin{aligned}
M_1 &: \theta_{\text{class}} > \theta_{\text{education}} > \theta_{\text{income}} > 0, \\
M_2 &: \theta_{\text{education}} > (\theta_{\text{class}}, \theta_{\text{income}}) > 0, \\
M_3 &: \theta_{\text{class}} = \theta_{\text{education}} = \theta_{\text{income}} > 0, \\
M_4 &: \text{"neither } M_1, M_2, \text{nor } M_3\text{"},
\end{aligned} \quad (2)$$

where θ_{class}, $\theta_{education}$, and θ_{income} denote the effects of social class, educational level, and income on attitude toward immigrants, respectively. Consequently, the goal is to quantify the evidence in the data for these three models to determine which model receives the most support.

Note that nuisance parameters (e.g., effects of control variables) are omitted in the above formulation of the models of interest to simplify the notation. Further note that additional competing constrained models could be formulated in this context as well. For the current application, however, we restrict ourselves to these models.

Application 2: The Importance of Postmaterialism for Young, Middle, and Old Generations

Experiences in preadult years are known to have a crucial impact on the development of basic values in later life. Due to the increase in welfare in recent decades, Generational Replacement Theory predicts that the values of younger generations are different from those of older generations. In particular, postmaterialistic values, such as the desire for freedom, self-expression, and quality of life, are expected to have increased for younger generations as a result of improved economic standards in Western countries (Inglehart and Abramson 1999; Welzel and Inglehart 2005).

In the European Values Study, generation was operationalized using an ordinal variable with three categories corresponding to a young, middle, or old generation. Similarly, postmaterialism has been measured on an ordinal scale as well, having three categories. When setting the younger generation as the reference group and using dummy variables for the middle and older generations, Generational Replacement Theory can be translated to an order-constrained model (M_1). We contrast it with a model that assumes no generation effect on postmaterialism (M_0) and with a complementary model that assumes neither an increased effect nor a zero effect (M_2).

The models of interest can be summarized as follows:

$$\begin{aligned} &M_0 : \theta_{old} = \theta_{middle} = 0, \\ &M_1 : \theta_{old} < \theta_{middle} < 0, \\ &M_2 : \text{``neither } M_0, \text{nor } M_1 \text{.''} \end{aligned} \qquad (3)$$

Furthermore, we hypothesize that the inclusion of order constraints on the generational effects of interest results in an increase of statistical power in comparison to testing the classical alternative, say, $M_3 : \theta_{young} \neq \theta_{middle} \neq 0$ versus the null model $M_2 : \theta_{young} = \theta_{middle} = 0$. In terms of the BIC, this

implies we would obtain more evidence against M_0 when testing it against the order-constrained model M_1 (if the constraints are supported by the data) than when testing M_0 against the unconstrained alternative M_3.

BIC Approximations of the Marginal Likelihood

In this section, extensions of the BIC are derived for a model with order (or inequality) constraints on certain model parameters. Consider an order-constrained model M_1 with d unknown model parameters, denoted by $\boldsymbol{\theta}$, which are restricted by r_1 order constraints, that is, $M_1 : \mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1$ where $[\mathbf{R}_1|\mathbf{r}_1]$ is an augmented $r_1 \times (d + 1)$ matrix containing the coefficients of the order constraints under M_1.

For example, the order-constrained model $M_1 : \theta_{\text{class}} > \theta_{\text{education}} > \theta_{\text{income}} > 0$ in equation (2) can be translated to

$$\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1 \Leftrightarrow \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_{\text{class}} \\ \theta_{\text{education}} \\ \theta_{\text{income}} \\ \theta_{\text{gender}} \\ \sigma^2 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad (4)$$

where the first element of $\boldsymbol{\theta}$ denotes the intercept, the fifth element the gender effect, and the sixth element denotes the error variance, which are nuisance parameters. The order-constrained model is nested in an unconstrained model that will be denoted by M_u.

The likelihood function under M_1 is a truncation of the likelihood under an unconstrained model, that is, $p_1(\mathbf{D}|\boldsymbol{\theta}) = p(\mathbf{D}|\boldsymbol{\theta}) \times I_{\Theta_1}(\boldsymbol{\theta})$, where $p(\mathbf{D}|\boldsymbol{\theta})$ denotes the likelihood function of the data $\mathbf{D}$ under the unconstrained parameter space Θ, Θ_1 denotes the order-constrained subspace under M_1, and I denotes the indicator function Θ. The prior for $\boldsymbol{\theta}$ under M_1 will be denoted by $p_1(\boldsymbol{\theta})$. Two different types of priors will be considered for approximating the marginal likelihoods under M_1 and M_u.

Truncated Unit Information Prior

First, we assume that the unconstrained posterior mode, denoted by $\tilde{\boldsymbol{\theta}}_u$, falls in the inequality-constrained space of model M_1, that is, $\mathbf{R}_1\tilde{\boldsymbol{\theta}} > \mathbf{r}_1$. The BIC approximation of the marginal likelihood under the inequality-constrained model is then obtained using a second-order Taylor expansion of the logarithm of the integrand around the posterior mode. This

approximation introduces in an error that is $\mathcal{O}(n^{-1})$.[1] Let us define $g(\boldsymbol{\theta}) = \log p_1(\mathbf{D}|\boldsymbol{\theta}) + \log p_1(\boldsymbol{\theta})$. Then, the marginal likelihood can be derived by

$$
\begin{aligned}
\log p_1(\mathbf{D}) &= \log \int_{\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1} p_1(\mathbf{D}|\boldsymbol{\theta})p_1(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \log \int_{\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1} \exp\{g(\boldsymbol{\theta})\}d\boldsymbol{\theta} \\
&= \log \int_{\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1} \exp\left\{g(\tilde{\boldsymbol{\theta}}_u) + \frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u)'\mathbf{H}(\tilde{\boldsymbol{\theta}}_u)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u)\right\}d\boldsymbol{\theta} + \mathcal{O}(n^{-1}) \\
&= \log p_1(\mathbf{D}|\tilde{\boldsymbol{\theta}}_u) + \log p_1(\tilde{\boldsymbol{\theta}}_u) \\
&\quad + \log \int_{\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1} \exp\left\{\frac{1}{2}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u)'\mathbf{H}(\tilde{\boldsymbol{\theta}}_u)(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}_u)\right\}d\boldsymbol{\theta} + \mathcal{O}(n^{-1}) \\
&= \log p_1(\mathbf{D}|\tilde{\boldsymbol{\theta}}_u) + \log p_1(\tilde{\boldsymbol{\theta}}_u) + \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|-\mathbf{H}(\tilde{\boldsymbol{\theta}}_u)| \\
&\quad + \log \Pr(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|\mathbf{D}, \mathbf{M}_u) + \mathcal{O}(n^{-1}),
\end{aligned}
$$

where $\mathbf{H}(\tilde{\boldsymbol{\theta}}_u)$ denotes the Hessian matrix of second-order partial derivatives of $g(\boldsymbol{\theta})$ evaluated at $\tilde{\boldsymbol{\theta}}_u$.

Hence, the only difference with the original derivation is that the resulting approximation also includes the posterior probability that the order constraints of $\mathbf{M}_1$ hold under the larger unconstrained model $\mathbf{M}_u$. From large sample theory, the unconstrained posterior mode can be approximated with the unconstrained MLE, that is, $\tilde{\boldsymbol{\theta}}_u \approx \hat{\boldsymbol{\theta}}_u$, and $-\mathbf{H}(\tilde{\boldsymbol{\theta}}_u) \approx nI_E(\hat{\boldsymbol{\theta}}_u)$, where $I_E(\hat{\boldsymbol{\theta}}_u)$ is the expected Fisher information matrix of one observation (which can be obtained using standard statistical software). This introduces an additional approximation error of $\mathcal{O}(n^{-1/2})$. Subsequently, the approximated logarithm of the marginal likelihood is given by

$$
\begin{aligned}
\log p_1(\mathbf{D}) &= \log p_1(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) + \log p_1(\hat{\boldsymbol{\theta}}_u) + \frac{d}{2}\log(2\pi) - \frac{d}{2}\log(n) \\
&\quad - \frac{1}{2}\log|I_E(\hat{\boldsymbol{\theta}}_u)| + \log \Pr(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|\mathbf{D}, \mathbf{M}_u) + \mathcal{O}\left(n^{-\frac{1}{2}}\right).
\end{aligned}
\tag{5}
$$

As was pointed out by Raftery (1995), certain terms cancel out when plugging in the so-called unit information prior (see also Kass and Wasserman 1995). The unit information prior has a multivariate normal distribution with mean equal to the MLE and variance equal to the inverse of the expected

Fisher information matrix of one observation, that is, $p^{UI}(\boldsymbol{\theta}) = N(\hat{\boldsymbol{\theta}}_u, I_E(\hat{\boldsymbol{\theta}}_u)^{-1})$. Under the constrained model M_1, we propose using a truncated unit information prior, that is,

$$p_1^{UI}(\boldsymbol{\theta}) = p^{UI}(\boldsymbol{\theta}) \times I(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1) \times Pr^{UI}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u)^{-1}, \qquad (6)$$

where the prior probability serves as a normalization contant so that the truncated unit information prior integrates to one, that is,

$$Pr^{UI}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u) = \int_{\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1} p^{UI}(\boldsymbol{\theta}) d\theta.$$

Evaluating the logarithm of the unconstrained unit information prior at the unconstrained MLE yields $\log p^{UI}(\hat{\boldsymbol{\theta}}_u) = -\frac{d}{2}\log(2\pi) + \frac{1}{2}\log|I_E(\hat{\boldsymbol{\theta}}_u)|$, and therefore, equation (5) becomes

$$\log p_1(\mathbf{D}) = \log p_1(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) + -\frac{d}{2}\log(n) + \log Pr(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|\mathbf{D}, M_u)$$
$$-\log Pr^{UI}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u) + \mathcal{O}(n^{-1/2}). \qquad (7)$$

The corresponding order-constrained BIC is then obtained by multiplying the logarithm of the approximated marginal likelihood by -2 and ignoring the error term. This yields

$$OC\text{-}BIC(M_1) = -2\log p_1(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) + d\log(n) - 2\log Pr(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|\mathbf{D}, M_u)$$
$$+ 2\log Pr^{UI}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u), \qquad (8)$$

where the first two terms form the ordinary BIC of model M_1 without the order constraints (i.e., M_u), and the additional third and fourth term are used for the evaluation of the order constraints of M_1 within M_u.

Next, we consider the case where the unconstrained posterior mode does not lie in the inequality-constrained subspace of M_1, that is, $\mathbf{R}_1\tilde{\boldsymbol{\theta}} \not> \mathbf{r}_1$. In this case, the second-order Taylor expansion of $g(\boldsymbol{\theta})$ around the unconstrained posterior mode (or MLE) may not be a good approximation. The rationale is that the mode under M_1, which will have a nonzero gradient, will lie on the boundary space where $\mathbf{R}_1\boldsymbol{\theta} = \mathbf{r}_1$.[2]

Because of the exponential tails of the normal distribution, a first-order Taylor expansion of $g(\boldsymbol{\theta})$ at the posterior mode under M_1, denoted by $\tilde{\theta}_1$, seems more appropriate (Avramidi 2000). For example, let us consider a simple inequality-constrained model, $M_1 : \theta \geq 0$, and let the unconstrained
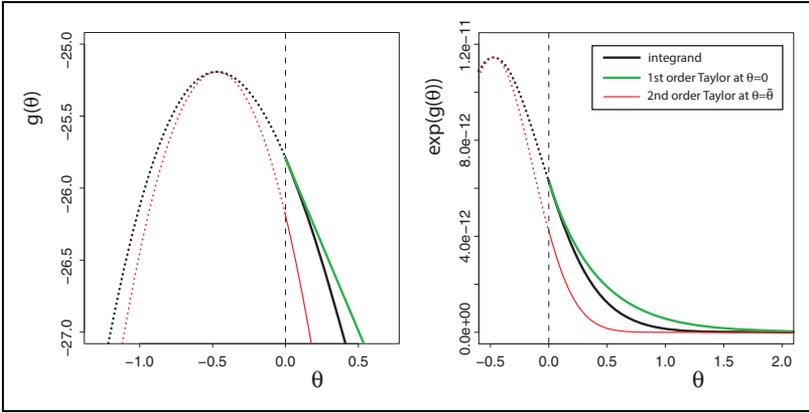
Figure 1. Plot of the log of prior times likelihood, $g(\theta)$ (black line), first-order Taylor approximation at $\theta = 0$ (green line), and second-order Taylor approximation around the unconstrained posterior mode of $\tilde{\theta} \approx -.5$ (red line), for an inequality-constrained model $M_1 : \theta \geq 0$. The left panel displays the functions of the log scale and the right panel on the regular scale. The inequality-constrained region under $M_1 : \theta \geq 0$ has solid lines, and the complement region has dotted lines.

mode be smaller than 0, that is, $\tilde{\theta} < 0$, so that the posterior mode under M_1 is located on the boundary, that is, $\tilde{\theta}_1 = 0$, which has a negative gradient, $g'(0) < 0$. The function $g(\theta)$ for such a situation is plotted in Figure 1 (black line; solid line under $\theta \geq 0$, dotted line under $\theta \not\geq 0$). The second-order Taylor approximation at the unconstrained mode is also plotted (red line; solid line under $\theta \geq 0$, dotted line under $\theta \not\geq 0$). A first-order Taylor expansion at $\tilde{\theta}_1 = 0$ can be used to approximate the function in the region $\theta \geq 0$ according to

$$g(\theta) = g(0) + g'(0)\theta + \mathcal{O}(\theta^2).$$

The marginal likelihood can then be approximated as follows:

$$\log p_1(\mathbf{D}) = \log \int_{\theta > 0} p_1(\mathbf{D}|\theta)p_1(\theta)d\theta = \log \int_{\theta > 0} \exp\{g(\theta)\}d\theta$$
$$\approx \log \int_{\theta > 0} \exp\{g(0) + g'(0)\theta\}d\theta = g(0) - \log(-g'(0)).$$

Hence, instead of the normal distribution which is used to compute the integral in the case of a second-order Taylor expansion, an exponential

distribution is used to compute the integral using this first-order Taylor expansion. The approximated line is also plotted in Figure 1 (green line).

The figure suggests that the second-order Taylor approximation at the unconstrained posterior mode is less accurate than the first-order Taylor approximation at the boundary point. This suggests that the approximated marginal likelihood under the inequality-constrained model will generally be better using first-order approximation at the boundary point in the case that the inequality constraints are not supported by the data. In the remainder of this article, however, we shall use the second-order Taylor approximation at the unconstrained posterior mode both when the posterior mode does and does not lie in the subspace of the inequality-constrained model under investigation.

When the order constraints are not supported by the data, the crudeness of the approximation is less important because the order-constrained model will not be selected because of the bad fit. Instead another, better fitting model will be selected for which the approximated marginal likelihood can be accurately estimated. Another reason for working with the second-order Taylor approximation is that it can easily be computed using equation (7), also for more complex systems of inequality constraints on multiple parameters, for example, $\theta_1 > \theta_2 > \theta_3 > 0$, than when using a first-order Taylor approximation at boundary point of the inequality-constrained subspace where the mode is located. The numerical experiments presented later illustrate that the approximation error is acceptable when the posterior mode does not lie in the constrained subspace.

It has been argued that data-based priors, such as the unit information prior, may result in Bayes factors that do not function as an Occam's razor when evaluating inequality-constrained models (Mulder 2014a, 2014b). To see that this is also the case for the unit information prior, the approximated Bayes factor of an inequality-constrained model M_1 against an unconstrained model M_u (where the inequality constraints are omitted) is given by

$$B_{1u}^{\mathrm{UI}} \approx \frac{\Pr(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|\mathbf{D}, M_u)}{\Pr^{\mathrm{UI}}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u)}.$$

This follows automatically from equation (7).

Now in the case of overwhelming evidence for M_1, that is, $\mathbf{R}_1\hat{\boldsymbol{\theta}}_u \gg \mathbf{r}_1$, both the posterior probability and the prior probability based on the unit information prior will be approximately 1, resulting in equal evidence for M_1 and M_u. This is a consequence of the fact that the unit information prior is concentrated around the MLE. Because both models fit the data equally well while the inequality-constrained model can be viewed as a less complex model

(because a "smaller" subspace is spanned). This property suggests that the approximated Bayes factor does not properly function as an Occam's razor.

Truncated Local Unit Information Prior

Due to the behavior of the unit information prior when evaluating order-constrained models, we consider a "local" unit information prior with a mean that is located on the boundary of the inequality-constrained space (we borrow the term "local" from Johnson and Rossell 2010). Note that the boundary space is equal to the parameter space under the null model $M_0 : \mathbf{R}_1\boldsymbol{\theta} = \mathbf{r}_1$. The rationale for centering the prior around the null space dates back at least to Jeffreys (1961) who argued that when the null model is false, the effects are expected to be close to the null; otherwise, there is no point in testing the null. This implies that the prior under the alternative model should be located around the null value. Furthermore, there have been reports in the literature where the use of such local priors results in desirable selection behavior when evaluating order-constrained models (e.g., Mulder 2014a; Mulder, Hoijtink, and Klugkist 2010). Here, we explore this class of priors for the BIC.

We set the mean of the local unit information prior equal to the MLE under the null model, denoted by $\hat{\boldsymbol{\theta}}_0$. Furthermore, the covariance matrix will be equal to the covariance matrix of the unit information prior. Thus, the unconstrained local unit information prior can be written as $p^{\mathrm{LUI}}(\boldsymbol{\theta}) = N(\hat{\boldsymbol{\theta}}_0, I_E(\hat{\boldsymbol{\theta}}_u)^{-1})$. The truncated prior under $M_1 : \mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1$ is then equal to

$$p_1^{\mathrm{LUI}}(\boldsymbol{\theta}) = p^{\mathrm{LUI}}(\boldsymbol{\theta}) \times I(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1) \times \frac{1}{\mathrm{Pr}^{\mathrm{LUI}}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u)}.$$

By applying equation (14) in Kass and Raftery (1995), changing the unit information prior to the local unit information prior results in an approximated logarithm of the marginal likelihood of

$$\log \hat{p}_1^{\mathrm{LUI}}(\mathbf{D}) = \log p(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) - \frac{d}{2}\log(n) - \frac{1}{2}(\hat{\boldsymbol{\theta}}_u - \hat{\boldsymbol{\theta}}_0)'I_E(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_u - \hat{\boldsymbol{\theta}}_0)$$
$$+ \log \mathrm{Pr}(\mathbf{R}\boldsymbol{\theta} > \mathbf{r}|\mathbf{D}, M_u) - \log Pr^{\mathrm{LUI}}(\mathbf{R}\boldsymbol{\theta} > \mathbf{r}|M_u). \tag{9}$$

Consequently, the approximated Bayes factor based on the local unit information prior of an inequality-constrained model against an unconstrained model is given by

$$B_{1u}^{\mathrm{LUI}} \approx \frac{\mathrm{Pr}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|\mathbf{D}, M_u)}{\mathrm{Pr}^{\mathrm{LUI}}(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u)}. \tag{10}$$

Now in the case of overwhelming evidence for M_1, in the sense that $\mathbf{R}_1\hat{\boldsymbol{\theta}}_u \gg \mathbf{r}_1$, the Bayes factor will be equal to the reciprocal of the prior probability that the inequality constraints hold under the unconstrained local unit information prior, that is, $B_{1u}^{\mathrm{LUI}} \approx (\mathrm{Pr}^L(\mathbf{R}_1\boldsymbol{\theta} > \mathbf{r}_1|M_u))^{-1}$, which is strictly larger than 1 because the prior mean is located on the boundary of the constrained space where $\mathbf{R}_1\boldsymbol{\theta} = \mathbf{r}_1$. Note that this prior probability can be viewed as a quantification of the relative size of the inequality-constrained subspace. For example, in the case of a diagonal covariance matrix, the prior probability of k one-sided constraints, $\boldsymbol{\theta} > \mathbf{0}$, is equal to 2^{-k}, and the prior probability of k order constraints, $\boldsymbol{\theta}_1 < \ldots < \boldsymbol{\theta}_k$, is equal to $(k!)^{-1}$, similar to the Bayes factors proposed by Mulder, Hoijtink, and Klugkist (2010) and Morey and Wagenmakers (2014).

Instead of working with equation (9), we consider a slightly cruder approximation where the third term, which quantifies prior fit, is omitted. This yields

$$
\begin{aligned}
\log \hat{p}_1^{\mathrm{LUI}*}(\mathbf{D}) = {} & \log p_u(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) - \frac{d}{2}\log(n) + \log \mathrm{Pr}(\mathbf{R}\boldsymbol{\theta} > \mathbf{r}|\mathbf{D}, M_u) \\
& - \log \mathrm{Pr}^{\mathrm{LUI}}(\mathbf{R}\boldsymbol{\theta} > \mathbf{r}|M_u).
\end{aligned} \tag{11}
$$

The rationale for omitting this term is that we are not interested in quantifying prior misfit. Another reason is that equation (11) can be combined with the ordinary BIC approximation for an unconstrained model (i.e., "$\log p(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) - \frac{d}{2}\log(n)$") to obtain the approximated Bayes factor in equation (10).

The terms on the right-hand side of equation (11) have the following intuitive interpretations. The first and second term can be interpreted as measures of model fit and model complexity of the unconstrained model where the inequality constraints are excluded (similar to the ordinary BIC approximation based on the unit information prior). The third term, which is the approximated posterior probability that the inequality constraints hold under the unconstrained model, can be interpreted as a measure of the relative fit of an order-constrained model M_1 relative to the unconstrained model M_u. Finally, the fourth term, which is the local prior probability that the order constraints hold under the unconstrained model, can be interpreted as a measure of the relative complexity of the order-constrained model M_1 relative to the unconstrained model M_u.

Thus, equation (11) will behave as an Occam's razor when evaluating order-constrained models by balancing the fit and complexity of the

order-constrained model. The corresponding order-constrained BIC based on the local unit information prior then yields

$$\text{OC-BIC}(M_1) = -2 \log p_u(\mathbf{D}|\hat{\boldsymbol{\theta}}_u) + d \log(n) - 2 \log \Pr(\mathbf{R}\boldsymbol{\theta} > \mathbf{r}|\mathbf{D}, M_u)$$
$$+ 2 \log \Pr^{\text{LUI}}(\mathbf{R}\boldsymbol{\theta} > \mathbf{r}|M_u).$$

(12)

Comparison With Other BIC Extensions

The order-constrained BIC in equation (12) shows some similarities with the BIC extensions proposed by Romeijn et al. (2012) and Morey and Wagenmakers (2014). In the proposal of Romeijn et al., the prior can be chosen by users allowing a subjective quantification of the relative size of the constrained space. Although this may be useful in certain situations, the BIC is typically used in an automatic fashion, and thus, it may be preferable to also let the prior probability be based on a default prior. The advantage of using the local unit information prior for this purpose is that it results in a reasonable default measure for the relative size of an order-constrained parameter space because the prior is centered on the boundary of the constrained space (unlike the [nonlocal] unit information prior). For example, when considering a univariate one-sided constraint, $\theta < 0$, the prior probability based on the local unit information prior will be $\frac{1}{2}$, which seems reasonable because half of the unconstrained space of θ is covered by the one-sided constraint.

Furthermore, Romeijn et al. set the posterior probability that the order constraints hold to 1 in the case the MLE is in agreement with the constraints and 0 elsewhere. This additional approximation step follows directly from large sample theory: When the sample size goes to infinity, the posterior probability converges to 1 if the true parameter value is an interior point of the order-constrained subspace and 0 if it is an interior point of the complement of this subspace. Thus, for extremely large samples, the prior-adapted BIC of Romeijn et al. may perform similarly to the order-constrained BIC in equation (12).

For small to moderately sized samples, however, setting the posterior probability to either 1 or 0 may result in crude approximations of the posterior probability. As will be shown in the empirical application in Application 1: Assessing the Importance of Different Dimensions of Socioeconomic Status subsection, for example, the posterior probabilities that two competing sets of order constraints hold under an unconstrained model are equal to

.50 and .18. Setting these probabilities to 1 and 0, respectively, would result in an unnecessarily crude estimate of the marginal likelihood. Instead, we recommend using the actual posterior probability that the order constraints hold based on the unconstrained approximated posterior (the third term in equation [12]).

In the proposal of Morey and Wagenmakers (2014), the prior probability that a specific ordering of d parameters hold, for example, $\theta_1 < \ldots < \theta_d$, is set to $1/d!$. This probability is thus based on the assumption that each ordering is equally likely a priori, similar to the priors proposed by Mulder et al. (2010) and Klugkist, Laudy, and Hoijtink (2005) when using Bayes factors. This probability however holds only for specific covariance structures such as a diagonal covariance structure. The prior probability may not be invariant for reparameterizations of the model (see also Mulder 2014a). For example, if we would define $\xi_{d'} = \theta_{d'} - \theta_{d'-1}$, for $d' = 2, \ldots, d$, and $\xi_1 = \theta_1$, the above order constraints would be equivalent to the one-sided constraints $(\xi_1, \ldots, \xi_{d-1}) > \mathbf{0}$. If one would use a prior diagonal covariance structure for ξ and zero means, the prior probability would be equal to $1/2^{d-1}$. This may be very different from $1/d!$, resulting in a serious violation of invariance to reparamaterizations. The prior probability based on the local unit information (the fourth term on the right-hand side of equation [12]) on the other hand would be invariant for such reparameterizations as the prior covariance structure is automatically transformed along with the reparameterization.

Numerical Analyses

The behavior of approximated Bayes factors based on the unit information prior and the local unit information prior will be investigated in a numerical example of the linear regression model, $y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \varepsilon_i$, with $\varepsilon_i \sim N(0, \sigma^2)$, for $i = 1, \ldots, n$. Here, θ_0 is the intercept, and θ_1 and θ_2 are the effects of the first and second predictor. We consider a model selection problem between an order-constrained model $M_1 : \theta_2 > \theta_1 > 0$, a null model $M_0 : \theta_2 = \theta_1 = 0$, and the complement model, $M_2 : \theta_2 \not> \theta_1 \not> 0$. To gain more insight into the behavior of the criterion as an Occam's razor, we also test the order-constrained model M_1 against the unconstrained model, $M_u : (\theta_1, \theta_2) \in \mathbb{R}^2$.

Statistical Evidence for Order-constrained Models

To better understand how the approximated Bayes factors quantify statistical evidence for order-constrained models, we computed the approximated

Bayes factors for data with $(\hat{\theta}_1, \hat{\theta}_2) = (a, 2a)$, for $a \in (-1.5, 1.5)$, while fixing $\hat{\theta}_0 = 0$, $\hat{\sigma}^2 = 1$, $n = 20$, and $\mathbf{X}'\mathbf{X} = [n\,0\,0;\,0\,n\,n/2;\,0\,n/2\,n]$ (the exact choice of these fixed values did not qualitatively affect the results). Thus, there is evidence for M_1, M_0, and M_2 when $a > 0$, $a = 0$, and $a < 0$, respectively.

The logarithm of the approximated Bayes factors can be found in Figure 2. Based on the approximated Bayes factors of M_1 versus M_u (left panel), we can see that the evidence based on the unit information prior (dotted line) for M_1 against M_u starts to decrease for larger effects (for approximately $a = .3$ and larger), which seems counterintuitive. Eventually, the weight of evidence (i.e., the log Bayes factor) converges to 0. Thus, in the case of overwhelming evidence for an order-constrained model, we obtain equal evidence for an order-constrained model M_1 that is fully supported by the data and the "larger" unconstrained model M_u when using the unit information prior, even though M_1 is a simpler model. This suggests that the approximated Bayes faction based on the unit information prior does not work as an Occam's razor when evaluating order-constrained models.

The evidence for M_1 against M_u based on the local unit information prior (solid line) on the other hand increases as a function of a. Eventually, the weight of evidence converges to the reciprocal of the prior probability of $\theta_2 > \theta_1 > 0$ under the unconstrained model M_u, which is strictly larger than 0. Furthermore, the local unit information prior results in more evidence for a model that is supported by the data in comparison to the unit information prior when comparing model M_1 versus model M_2 (Figure 2, middle panel) and model M_1 versus M_0 (Figure 2, right panel). Based on these considerations, we conclude that the order-constrained BIC based on the local unit information prior better balances fit and complexity when evaluating order-constrained models than the order-constrained BIC based on the unit information prior.

Error Probabilities

Next, we investigate the probabilities of selecting the true data generating model when including order constraints in the alternative model or not. First, we consider testing the null model, $M_0 : \theta_1 = \theta_2 = 0$, against an unconstrained alternative, $M_u : \boldsymbol{\theta} \in \mathbb{R}^2$, using the ordinary BIC. Second, we consider testing the null model $M_0 : \theta_1 = \theta_2 = 0$ versus $M_1 : \theta_2 > \theta_1 > 0$ against two order-constrained alternative, namely, $M_2 : \theta_2 \not> \theta_1 \not> 0$, using the two order-constrained BICs. Note that the BIC for M_0, with no inequality
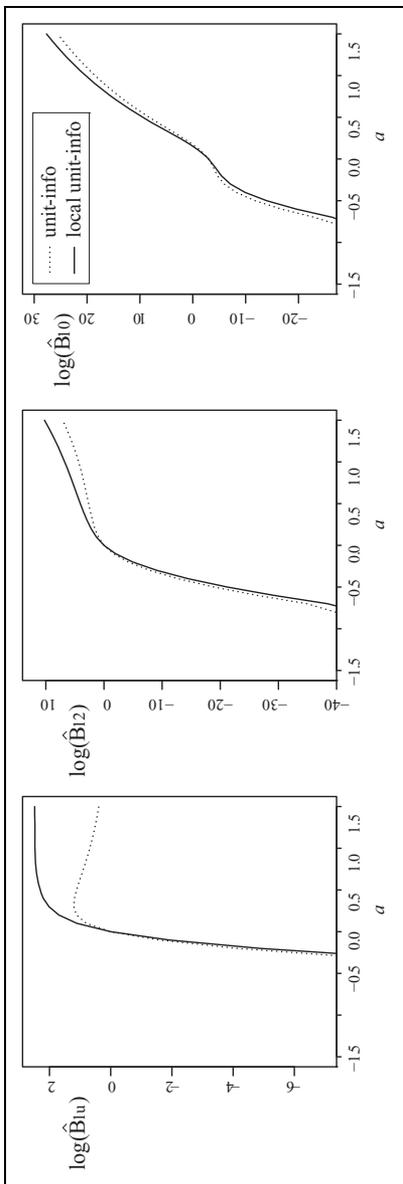
Figure 2. The logarithm of the approximated Bayes factors based on the unit information prior $\log(\hat{B}^{UI})$ (dotted line), and the local unit information prior $\log(\hat{B}^{L})$ (solid line) of $M_1 : \theta_2 > \theta_1 > 0$ versus $M_u : (\theta_1, \theta_2) \in R^2$ (left panel), of M_1 versus the complement model M_2 (middle panel), and of M_1 versus $M_0 : \theta_2 = \theta_1 = 0$ (right panel). The criteria are plotted for $n = 20$ as a function of a, where $(\hat{\theta}_1, \hat{\theta}_2) = (a, 2a)$.

17
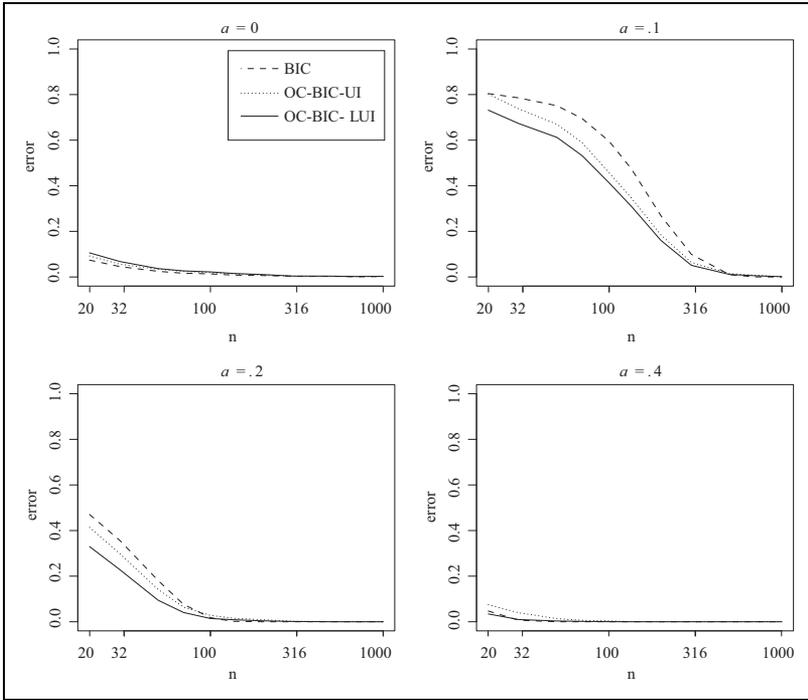
Figure 3. Probability of selecting the wrong model when using the ordinary Bayesian information criterion (BIC) for testing $M_0 : \theta_1 = \theta_2 = 0$ against $M_u : \boldsymbol{\theta} \in \mathbb{R}^2$ (dashed line), and the two order-constrained BICs when testing $M_0 : \theta_1 = \theta_2 = 0$, $M_1 : \theta_2 > \theta_1 > 0$, and $M_2 : \theta_2 \not> \theta_1 \not> 0$ (dotted and solid line for the nonlocal and local unit-information prior, respectively) for true effects of $(\theta_1, \theta_2) = (a, 2a)$, for $a = 0$, .1, .2, and .4. The sample size on the *x*-axis is on a logarithmic scale.

constraints, is the same in both tests. Further note that because the second test contains three models instead of two, the error probabilities in the second selection problem will be slightly larger when M_0 is true, as a result of the design. The true effects will be set to $(\theta_1, \theta_2) = (a, 2a)$, for $a = 0$, so that M_0 is true, and $a = .1$, .2, and .4, so that M_u (M_1) is true in the first (second) test.

Figure 3 displays the error probabilities as a function of the sample size (on a log scale). All the criteria show consistent behavior in the sense that the error probabilities go to 0 as the sample size grows. Furthermore, we see that when M_0 is true (upper left panel), the error probabilities are very similar and the ordinary BIC in test 1 results in the smallest errors. In the case of a true effect in the direction of the order constraints of M_1, we see that the
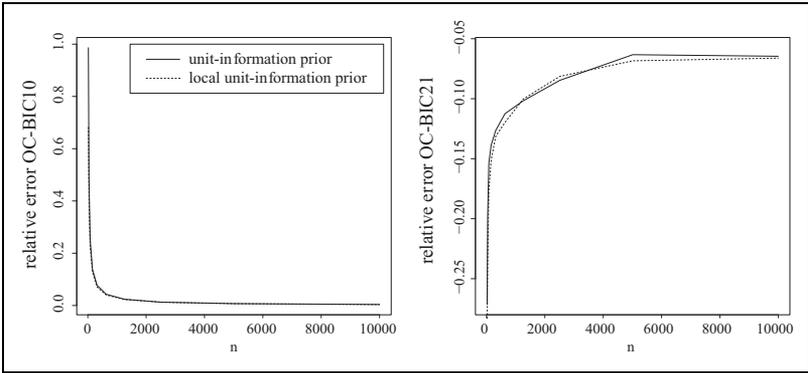
Figure 4. Left panel: Relative approximation error of the order-constrained Bayesian information criterion (BIC) of $M_1 : \theta_2 > \theta_1 > 0$ versus $M_0 : \theta_1 = \theta_2 = 0$ when $(\hat{\theta}_1, \hat{\theta}_2) = (.5, 1)$. Right panel: Relative approximation error of the order-constrained BIC of $M_2 : \theta_1 \not> \theta_2 \not> 0$ versus $M_1 : \theta_2 > \theta_1 > 0$ when $(\hat{\theta}_1, \hat{\theta}_2) = (-.5, -1)$.

order-constrained BIC based on the local unit-information prior results in considerably smaller errors than the other criteria.

The error probabilities of the order-constrained BIC based on the local unit-information prior were only slightly larger in the case of a nonzero effect. This is partly a consequence of the design of the test having three instead of two models under investigation. We conclude that overall, the order-constrained BIC based on the local unit-information prior performs best in terms of error probabilities.

Approximation Errors of the Order-constrained BICs

Finally, we investigated the relative approximation errors of the order-constrained BICs by comparing them to nonapproximated counterparts, for example, $\frac{\log B_{12} - \log \hat{B}_{12}}{\log B_{12}}$ for model M_1 against M_2. The approximation errors were investigated when the order-constrained model is supported by the data, namely, when testing $M_1 : \theta_2 > \theta_1 > 0$ versus $M_0 : \theta_1 = \theta_2 = 0$, with $(\hat{\theta}_1, \hat{\theta}_2) = (.5, 1)$, and when the order constrained is not supported by the data, namely, when testing $M_1 : \theta_2 > \theta_1 > 0$ against its complement $M_2 : \theta_1 \not> \theta_2 \not> 0$, with $(\hat{\theta}_1, \hat{\theta}_2) = (-.5, -1)$, while increasing the sample size.

The results can be found in Figure 4. As can be seen from the left panel, the relative error goes to 0 fast when the effects are in agreement with the

order constraints of model M_1. When the effect are not in agreement with the constraints (right panel), we see that the relative error does not go to zero. This is a consequence of the somewhat crude approximation we already observed in Figure 1 (red line). The approximation error, however, is not large enough to be a serious practical problem. Other settings resulted in qualitatively similar results.

Software

The R package "`BICpack`" was developed for evaluating order-constrained models using the order-constrained BIC based on the local unit-information prior. The R functions can be downloaded from `www.github.com/jomulder/BICpack`[3] or from CRAN in the near future. The order-constrained BIC based on the truncated unit information prior was not considered because of its poorer performance that we observed in the numerical simulations. The package makes use of the `mvtnorm`-package (Genz et al. 2016) for computing the probabilities in equation (11). The key function is "`bic_oc`," which can be used for computing the order-constrained BIC for various statistical models including generalized linear models and survival models. As input the function needs a fitted model object (e.g., a fitted `glm`-object or `coxph`-object), a character string denoting the order constraints on certain parameters, and a Boolean argument denoting whether the order-constrained subspace or its complement is considered (the default is the order-constrained subspace).

For example, in the case of a regression model with three predictors, say, `X1`, `X2`, and `X3`, on an outcome variable `y`, and it is expected that `X1` has the largest effect on the outcome variable, followed by `X2`, and `X3` is expected to have the smallest effect, and all effects are expected to be positive, the order-constrained BIC can be computed by executing the following lines

```
fit1 <- glm(y~X1 + X2 + X3, data)
bic_oc(fit1, "0 < X1 < X2 < X3")
```

The use of the function will be illustrated in two empirical applications in the next section.

Empirical Applications Revisited

The models from the applications in the second section are evaluated using the order-constrained BIC based on the local unit information prior using the

R package `BICpack`. For the empirical analyses presented in this section, only the European Values Study data of Germany are considered.

Application 1: Assessing the Importance of Different Dimensions of Socioeconomic Status

Model 1 can be fitted in R using the `lm` function:

```
lm1 <- lm(atti_immi˜class + education + income + gender,
          data=EVS_Germany)
```

The estimated coefficients of interest were $(\beta_{\text{class}}, \beta_{\text{education}}, \beta_{\text{income}}) = (.312, .250, .041)$, with standard errors .067, .075, and .072, respectively.

The order-constrained BIC for model $M_1 : \theta_{\text{class}} > \theta_{\text{education}} > \theta_{\text{income}} > 0$ in equation (2) can then be computed using the new `bic_oc`-function:

```
bic_oc(lm1, "class > education > income > 0")
```

This resulted in a BIC of 3,918.46. The function also provides the posterior probability that the constraints hold under the unconstrained model, which was equal to .50. Next, the BIC for model $M_2 : \theta_{\text{education}} > (\theta_{\text{class}}, \theta_{\text{income}}) > 0$ is computed using the command

```
bic_oc(lm1, "education > (class, income) > 0")
```

The resulting order-constrained BIC was 3,921.98. For this set of constraints, the posterior probability under the unconstrained model equaled .18.

The BIC for model $M_3 : \theta_{\text{class}} = \theta_{\text{education}} = \theta_{\text{income}} > 0$ is computed. First, the model is fitted with the equality constraints on the effect but without the inequality constraint. Because the effects of social class, education, and income are equal under M_3, the regression model in equation (1) becomes

$$\text{atti\_immi} = \theta_0 + (\text{class} + \text{education} + \text{income}) \times \theta_{\text{class.educ.income}} + \text{error}$$

where $\theta_{\text{class.educ.income}}$ denotes the equal effect of social class, educational level, and income on attitude toward immigrants. Thus, this model can be fitted by including the sum of the class, education, and income as a linear predictor:

```
EVS_Germany$class.educ.income <- EVS_Germany$class +
    EVS_Germany$education + EVS_Germany$income
```

Table 1. Order-constrained BICs and Posterior Model Probabilities for the Competing Models in Application 1.

Models	OC-BIC*	$P(M_t|\mathbf{D})$
$M_1 : \theta_{class} > \theta_{education} > \theta_{income} > 0$	3,918.46	.391
$M_2 : \theta_{education} > (\theta_{class}, \theta_{income}) > 0$	3,921.98	.067
$M_3 : \theta_{class} = \theta_{education} = \theta_{income} > 0$	3,917.84	.533
$M_4 :$ "neither M_1, M_2, nor M_3"	3,926.13	.008

Note: BIC = Bayesian information criterion.

```
lm2  <-  lm(atti_immi ˜ class.educ.income  +  gender,
    data=EVS_Germany)
```

The order-constrained BIC can then be computed based on the resulting fitted model:

```
bic_oc(lm2,"class.educ.income > 0")
```

This resulted in a BIC of 3,917.84.

Finally, to compute the BIC of the complement model $M_4 :$ "neither M_1, M_2, nor M_3, ", first note that the marginal likelihood of the union of M_1, M_2, and M_3 would be the same as the marginal likelihood of the union of only M_1 and M_2 because M_3 has zero probability due to the presence of the equality constraints of M_3. Thus, we need to compute the marginal likelihood of the complement model of the joint of models M_1 and M_2. First, we combine the two sets of order constraints in one vector and then compute the order-constrained BIC using the new function:

```
constraints_M4 <- c("class > education > income > 0",
                    "education > (class, income) > 0")
bic_oc(lm1, constraints_M4, complement = TRUE)
```

This resulted in a BIC of 3,926.13. The BIC values are summarized in Table 1. From these values, we can conclude that model M_3 receives most support, but the evidence is negligible in comparison to the evidence for the order-constrained model M_1, given the BIC difference of .62.[4] The evidence for M_2 and M_3 is considerably lower than for M_1 and M_3.

For interpretation purposes, it can be useful to translate the BICs to posterior model probabilities. A posterior model probability quantifies the probability of the data having been generated by one of the models considered, after observing the data given certain prior model probabilities. This probability is conditional on the data having been generated by one of the models considered.

In this application, we assume equal prior probabilities for the models. The posterior model probabilities can be computed from the BIC values using the "`postprob`" function in `BICpack`. The posterior probabilities together with the BICs can be found in Table 1. Hence, the posterior probability for model M_3, which assumes equal and positive effects of social class, education, and income on attitude toward immigrants, is largest with 53.3 percent. The posterior probability of M_1, which assumed ordered positive effects of social class, education, and income based on the Ethnic Competition Theory, is only slightly smaller with 39.1 percent. There is not much evidence for either M_2 or M_2, given their posterior probabilities of 6.7 percent and 0.8 percent, respectively. In sum, we can conclude that there is considerable model uncertainty, and more data would be needed to choose a single best model.

Application 2: The Importance of Postmaterialism for Young, Middle, and Old Generations

Because the outcome variable "postmaterialism" has an ordinal measurement level with three categories ("low," "medium," and "high"), an ordinal regression model can be fitted using the `polr`-function of the `MASS`-package. Thus, the ordinal variable "postmaterialism" is regressed to the ordinal predictor "generation" with categories "young," "middle," and "old" while controlling for "gender," "income," and "education":

```
fit3 <- polr(postmaterial~generation + gender + income +
             education, data=EVS_Germany, Hess=TRUE)
```

In the fitted model, the "young" generation is the reference group, and dummy variables are created for the "middle" and "old" generation. These variables are called "`generationmiddle`" and "`generationold`" in the fitted `polr`-object. The estimated effects under this model were equal to $(\hat{\theta}_{\text{generationmiddle}}, \hat{\theta}_{\text{generationold}}) = (-.444, -.848)$, having standard errors of .154 and .150, respectively.

Thus, the order-constrained BIC of model $M_1 : \theta_{\text{generationold}} < \theta_{\text{generationmiddle}} < 0$, representing the Generational Replacement Theory, can be computed by the command

```
bic_oc(fit3, "generationold < generationmiddle < 0")
```

The resulting BIC equaled 3,154.82.

Next, the BIC of the null model $M_0 : \theta_{\text{generationold}} = \theta_{\text{generationmiddle}} = 0$ is computed with no generation effect. Because this model does not

Table 2. Order-constrained BICs and Posterior Model Probabilities for the Competing Models in Application 2.

Models	OC-BIC*	$P(M_t|\mathbf{D})$
$M_0 : \theta_{old} = \theta_{middle} = 0$	3,177.69	0.00
$M_1 : \theta_{old} < \theta_{middle} < 0$	3,154.82	1.00
$M_2 :$ "neither M_0, nor M_1"	3,170.15	0.00

Note: BIC = Bayesian information criterion.

contain any order constraints, we can simply compute an ordinary BIC. This can also be done using the `bic_oc`-function by omitting any order constraints:

```
fit4 <- polr(postmaterial˜1 + gender + income +
         education, data=EVS_Germany, Hess=TRUE)
bic_oc(fit4)
```

The resulting BIC was equal to 3,177.69.

Finally, the BIC of the complement model was computed. Similarly to the previous example, this can be done as follows:

```
bic_oc(fit3, "generationold < generationmiddle < 0",
complement=TRUE)
```

This resulted in a BIC of 3,170.15. The BICs and respective posterior model probabilities can be found in Table 2. Clearly, there is overwhelming evidence for M_1 which implies that postmaterialism has increased for younger generations.

Finally, we show that the inclusion of order constraints in the alternative model results in more evidence against a null model if the order constraints are supported by the data. First, note that the BIC for the order-constrained model, M_1, against the null model, M_0, equals $BIC(M_1, M_0) = BIC(M_1) - BIC(M_0) = 3,154.82 - 3,177.69 = -22.87$. The BIC for an unconstrained alternative model, $M_3 : \theta_{generationold} \neq \theta_{generationmiddle} \neq 0$, against the null model equals $BIC(M_3, M_0) = BIC(M_3) - BIC(M_0) = 3,158.18 - 3,177.69 = -19.51$.[5] Hence, the inclusion of order constraints results in a substantial increase of the evidence against the null model in the case where the order constraints are supported by the data. We also get a more informative answer about how the effects are related to each other in the case there is evidence against the null model than when testing the null against an unconstrained alternative.

Discussion

We have presented two extensions of the BIC for evaluating models with order constraints on certain parameters of interest. In the first extension, a truncated unit information prior was considered under the order-constrained model, and in the second extension, a truncated local unit information prior was considered. Theoretical considerations and numerical analyses revealed that the local unit information prior resulted in better model selection behavior than the nonlocal unit information prior for order-constrained model selection.

The new order-constrained BIC based on the local unit-information prior can easily be computed using the new R package "BICpack." This will allow researchers to test multiple social theories that can be translated into conflicting sets of equality and order constraints on the parameters of interest. The methodology can also be used for testing directed effects of ordinal predictors, as these expectations can be translated into order-constrained models in a natural manner.

ORCID iD

J. Mulder ⓘ https://orcid.org/0000-0001-7007-3086

Notes

1. A quantity being $\mathcal{O}(n^{-1})$ implies that $n\mathcal{O}(n^{-1})$ converges to a constant as $n \to \infty$.
2. In the case $\mathbf{R}_1\tilde{\boldsymbol{\theta}} = \mathbf{r}_1$, the second-order expansion may still be appropriate.

3. Run "`devtools:: install_github("jomulder/BICpack")`" in R to install the BICpack package.
4. Typically a Bayesian information criterion (BIC) difference of 10 points is needed in order to rule out a model.
5. The BIC for M_3 can be obtained by running "`bic_oc(fit3)`."

References

Avramidi, I. 2000. *Lecture Notes Methods of Mathematical Physics MATH 536*. New Mexico Institute of Mining and Technology. Retrieved November 15, 2019 (https://www.researchgate.net/publication/266229957_Lecture_Notes_Methods_ of_Mathematical_Physics_MATH_536).

Böing-Messing, F., M. van Assen, A. Hofman, H. Hoijtink, and J. Mulder. 2017. "Bayesian Evaluation of Constrained Hypotheses on Variances of Multiple Independent Groups." *Psychological Methods* 22:262-87.

Braeken, J., J. Mulder, and S. Wood. 2015. "Relative Effects at Work: Bayes Factors for Order Hypotheses." *Journal of Management* 41:544-73.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Cohen, A. K., M. Rai, D. H. Rehkopf, and B. Abramsn. 2013. "Educational Attainment and Obesity: A Systematic Review." *Obesity Review* 14:989-1005.

de Jong, J., T. Rigotti, and J. Mulder. 2017. "One after the Other: Effects of Sequence Patterns of Breached and Overfulfilled Obligations." *European Journal of Work and Organizational Psychology* 26:337-55.

Genz, A.F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl . . . Hothorn, T. 2016. *R-Package 'mvtnorm' [Computer Software Manual]* (R package version 1.14. 4—For new features, see the 'Changelog' file (in the package source)).

Hoijtink, H. 2011. *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. New York: Chapman & Hall/CRC.

Inglehart, R. and P. R. Abramson. 1999. "Measuring Postmaterialism." *American Political Science Review* 93:665-77.

Jeffreys, H. 1961. *Theory of Probability*. 3rd ed. New York: Oxford University Press.

Johnson, V. E. and D. Rossell. 2010. "On the Use of Non-local Prior Densities in Bayesian Hypothesis Tests." *Journal of the Royal Statistical Society Series B* 72:143-70.

Kass, R. E. and A. E. Raftery. 1995. "Bayes Factors." *Journal of American Statistical Association* 90:773-95.

Kass, R. E. and L. Wasserman. 1995. "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion." *Journal of the American Statistical Association* 90:928-34.

Klugkist, I., O. Laudy, and H. Hoijtink. 2005. "Inequality Constrained Analysis of Variance: A Bayesian Approach." *Psychological Methods* 10:477-93.

Kluytmans, A., R. V. D. Schoot, J. Mulder, and H. Hoijtink. 2012. "Illustrating Bayesian Evaluation of Informative Hypotheses for Regression Models." *Frontiers in Psychology* 3:Article ID 2.

Lee, S.-Y. and X.-Y. Song. 2007. "A Unified Maximum Likelihood Approach for Analyzing Structural Equation Models with Missing Nonstandard Data." *Sociological Methods & Research* 35:352-81.

Morey, R. D. and E.-J. Wagenmakers. 2014. "Simple Relation between Bayesian Order-restricted and Point-null Hypothesis Tests." *Statistics and Probability Letters* 92:121-24.

Mulder, J. 2014a. "Bayes Factors for Testing Inequality Constrained Hypotheses: Issues with Prior Specification." *British Journal of Statistical and Mathematical Psychology* 67:153-71.

Mulder, J. 2014b. "Prior Adjusted Default Bayes Factors for Testing (In)Equality Constrained Hypotheses." *Computational Statistics and Data Analysis* 71: 448-63.

Mulder, J. and J.-P. Fox. 2019. "Bayes Factor Testing of Multiple Intraclass Correlations." *Bayesian Analysis* 14:521-52.

Mulder, J., H. Hoijtink, and I. Klugkist. 2010. "Equality and Inequality Constrained Multivariate Linear Models: Objective Model Selection Using Constrained Posterior Priors." *Journal of Statistical Planning and Inference* 140:887-906.

Mulder, J. and L. R. Pericchi. 2018. "The Matrix-F Prior for Estimating and Testing Covariance Matrices." *Bayesian Analysis* 13:1189-210.

Raftery, A. E. 1986. "Choosing Models for Cross-classifications." *American Sociological Review* 51:145-46.

Raftery, A. E. 1993. "Bayesian Model Selection in Structural Equation Models." Pp. 163-80 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Beverly Hill, CA: Sage.

Raftery, A. E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111-63.

Romeijn, J.-W., R. van de Schoot, and H. Hoijtink. 2012. "One Size Does Not Fit All: Proposal for a Prior-adapted BIC." Pp. 87-105 in *Probabilities, Laws, and Structures*, edited by D. Dieks, W. Gonzalez, S. Hartmann, M. St oltzner, and M. Weber. Dordrecht, the Netherlands: Springer.

Scheepers, P., M. Gijsberts, and M. Coenders. 2002. "Ethnic Exclusion in European Countries: Public Opposition to Civil Rights for Legal Migrants as a Response to Perceived Ethnic Threat." *European Sociological Review* 18:17-34.

Schwarz, G. E. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6: 461-64.

van de Schoot, R.J. Mulder, H. Hoijtink, M. Van Aken, J. Semon Dubas, B. Orobio de Castro . . . Romeijn, J.-W. 2012. "An Introduction to Bayesian Model Selection

for Evaluating Informative Hypotheses." *European Journal of Developmental Psychology* 8:713-29.

van der Waal, J., W. de Koster, and J. ten Kate. 2015. "Educational Attainment and Obesity in the Netherlands: Class or Status?" Paper presented at 'Dag van de Sociologie,' Amsterdam.

Vermunt, J. K. 1997. *Log-linear Models for Event Histories*. Thousand Oaks, CA: Sage.

Welzel, C. and R. Inglehart. 2005. "Liberalism, Postmaterialism, and the Growth of Freedom." *International Review of Sociology* 15:81-108.

Author Biographies

J. Mulder is an associate professor in Bayesian statistical methods at Tilburg University. Also, Jheronimus Academy of Data Science, 's Hertogenbosch, the Netherlands. His work currently focusses on Bayesian model selection and relational event history modeling in the social and behavioral sciences.

A. E. Raftery is the Boeing International Professor of Statistics and Sociology at the University of Washington. He works on the development of new statistical methods for the social, environmental and health sciences. An elected member of the U.S. National Academy of Sciences, he was identified as the world's most cited researcher in mathematics for the decade 1995-2005 by Thomson-ISI.