OBJECTIVE BAYESIAN COMPARISON OF CONSTRAINED ANALYSIS OF VARIANCE MODELS

Guido Consonni and Roberta Paroli
Università Cattolica del Sacro Cuore

In the social sciences we are often interested in comparing models specified by parametric equality or inequality constraints. For instance, when examining three group means \( \mu_1, \mu_2, \mu_3 \) through an analysis of variance (ANOVA), a model may specify that \( \mu_1 < \mu_2 < \mu_3 \), while another one may state that \( \{ \mu_1 = \mu_3 \} < \mu_2 \), and finally a third model may instead suggest that all means are unrestricted. This is a challenging problem, because it involves a combination of nonnested models, as well as nested models having the same dimension. We adopt an objective Bayesian approach, requiring no prior specification from the user, and derive the posterior probability of each model under consideration. Our method is based on the intrinsic prior methodology, suitably modified to accommodate equality and inequality constraints. Focussing on normal ANOVA models, a comparative assessment is carried out through simulation studies. We also present an application to real data collected in a psychological experiment.

Key words: ANOVA, Bayes factor, Bayesian model choice, hypothesis testing, inequality constraint, intrinsic prior.

1. Introduction

Constrained statistical inference has a long history in the statistical literature; see the classic monograph Barl, Barlow, Bartholomew, Brenner, and Brunk (1972), followed by Robertson, Wright, and Dykstra (1988) and the more recent book Silvapulle and Sen (2005). Models specified by parameter constraints are common in the social sciences, and especially in psychology; see Klugkist, Laudy, and Hoijtink (2005) and Hoijtink, Klugkist, and Boelen (2008). For instance, consider a three-way normal ANOVA with group means \( \mu_j \). One possible model is \( M_1 : \mu_1 < \mu_2 < \mu_3 \), while another one is \( M_2 : \{ \mu_1 = \mu_3 \} < \mu_2 \). Two special models stand out: the encompassing, or unconstrained, model \( M_e : \mu_1, \mu_2, \mu_3 \), wherein no constraint is imposed on the parameters, and the null model \( M_0 : \mu_1 = \mu_2 = \mu_3 \). Our goal is to compare these models.

We use the Bayes factor (BF) (Kass & Raftery, 1995) as a criterion for testing between two hypotheses or models, and more generally for model selection. Relative to more conventional measures of evidence in testing scenarios, such as the \( p \)-value, the BF exhibits appealing features in applied statistical research; see Wagenmakers (2007) for a review with emphasis on psychological studies, and Johnson (2013) for a critical appraisal together with suggestions to revise current thresholds of evidence. In particular, Bayes factors have been shown to be especially effective when testing constrained hypotheses; see Klugkist and Hoijtink (2007) and Mulder (2014a) for further discussion. The computation of the Bayes factor requires the specification of a separate prior distribution on the parameters of each model. For a fully Bayesian model comparison, one should add a prior distribution on the space of models probabilities in order to obtain the posterior probability for the set of models under consideration, which entirely summarizes our inference; see O’Hagan and Forster (2004, Ch. 7).

Electronic supplementary material The online version of this article (doi:10.1007/s11336-016-9516-y) contains supplementary material, which is available to authorized users.

Correspondence should be made to Guido Consonni, Università Cattolica del Sacro Cuore, Milan, Italy. Email: guido.consonni@unicatt.it; https://sites.google.com/site/consonnibayes/home

© 2016 The Psychometric Society
Consider for simplicity two sampling models for the observations \( y \), namely \( M_1 : \{ f_1(y|\theta_1 \in \Theta_1) \} \) and \( M_2 : \{ f_2(y|\theta_2 \in \Theta_2) \} \). Let \( p_1(\theta_1) \) and \( p_2(\theta_2) \) be the priors on the parameters under each of the two models. The marginal likelihood for model \( M_i \) is \( m_i(y) = \int_{\Theta_i} f_i(y|\theta_i) p_i(\theta_i) d\theta_i \), \( i = 1, 2 \). The Bayes factor of model \( M_1 \) against model \( M_2 \) for given data \( y \) is the ratio \( m_1(y) / m_2(y) \) of the two marginal likelihoods. Of particular interest is the case in which \( M_1 \) is nested into \( M_2 \), that is, the two sampling densities belong to the same family with \( \Theta_1 \subset \Theta_2 \). In this paper we consider general nested situations where the two parameter subsets may have different dimensions, as in the ANOVA models \( M_2 \) and \( M_e \) above, as well as the same dimension, as in the case of \( M_1 \) and \( M_e \).

Since the assignment of the parameter priors \( p_i(\theta_i) \) can be problematic, we take an objective Bayesian standpoint wherein default priors are used, without subjective inputs from the user; see Berger (2006). More specifically, we focus on objective Bayes model selection; for a comprehensive review see Pericchi (2005). In particular, we consider the intrinsic prior approach (Berger & Pericchi, 1996; Moreno, 1997; Pérez & Berger, 2002), which over the years has emerged as a powerful methodology for comparing nested models in a variety of settings; see for instance Casella and Moreno (2006); Girón, Martínez, Moreno, and Torres (2006); Consonni and La Rocca (2008); and Leon-Novelo, Moreno, and Casella (2012). Casella, Girón, Martínez, and Moreno (2009) and Moreno, Girón, and Casella (2010) study consistency in model selection based on intrinsic priors, namely whether the sequence of posterior probabilities of the assumed true model converges to one as the sample size grows.

Klugkist et al. (2005), Klugkist and Hoijtink (2007), and Laudy and Hoijtink (2007) have introduced a methodology, named encompassing prior, which deals specifically with the comparison of inequality constrained models. For a critical discussion, see Stern (2005). More recently, objective Bayesian methods for the comparison of inequality constrained models have been presented; they include Mulder, Hoijtink, and Klugkist (2010), Wesel, Hoijtink, and Klugkist (2011), Hoijtink (2013), and Mulder (2014c) (and further references therein). In this paper we propose an objective Bayesian comparison of constrained models which is (i) principled, because it is based on the well-established theory of intrinsic priors for model selection; (ii) fully automatic, i.e., it does not involve parametric fine-tuning; (iii) can deal simultaneously with inequality and equality constraints (the latter being treated exactly and not only approximately); and (iv) does not make use of empirical training samples. The starting point of our approach is broadly similar to that described in Pericchi, Liu, and Núñez Torres (2008), although our analytical development is quite distinct; see our Discussion at the end of the paper for further details.

The rest of the paper is organized as follows: Section 2 presents the general framework of our methodology, which is implemented for one-way ANOVA models presented in Sect. 3. Section 4 presents simulations and an application. Finally, Sect. 5 offers several points for discussion.

2. Objective Bayesian Comparison of Constrained Models

In this section we develop a general methodology for the objective Bayesian comparison of models with inequality/equality constraints.

2.1. Intrinsic Priors

Consider two sampling models for the observations \( y \), namely \( M_1 : \{ f_1(y|\theta_1) \} \) and \( M_2 : \{ f_2(y|\theta_2) \} \). Let \( p_1^N(\theta_1) \) and \( p_2^N(\theta_2) \) be estimation-based priors (e.g., reference priors, or other conventional priors; here the superscript “\( N \)” stands for noninformative). There are two reasons why such priors are not suitable for testing or model choice: (i) they are typically improper, and (ii) each prior is exclusively based on its own model, and thus the two priors are not “linked.” Point
that is less known, but equally crucial, and is related to compatibility of priors across models; see Consonni and Veronese (2008) for some general discussion, and Consonni and La Rocca (2008) and Casella and Moreno (2009) with specific reference to intrinsic priors.

To deal with (i) partial BFs were first introduced followed by a more robust version, namely intrinsic BF (IBF); see Berger and Pericchi (1996). The IBF is asymptotically equivalent to an actual BF computed using a pair of intrinsic priors (one under each model). If \( M_1 \) is nested into, and of lower dimension than \( M_2 \), the intrinsic prior for \( \theta_1 \) coincides with the original prior, i.e., \( p_1^1(\theta_1) = p_1^N(\theta_1) \). On the other hand, the intrinsic prior for \( \theta_2 \) can be constructed in two steps:

(i) Conditional intrinsic prior (CIP)

\[
p_2^1(\theta_2|\theta_1) = p_2^N(\theta_2) E_{\theta_1}^{M_2} \left( \frac{f_1(x|\theta_1)}{m_2^N(x)} \right)
\]

(ii) Intrinsic prior (IP)

\[
p_2^1(\theta_2) = \int p_2^1(\theta_2|\theta_1) p_1^N(\theta_1) d\theta_1,
\]

where \( m_2^N(x) = \int f_2(x|\theta_2) p_2^N(\theta_2) d\theta_2 \), and the expectation appearing in (1) is with respect to the sampling distribution under \( M_2 \), \( f_2(x|\theta_2) \), with \( x \) a random vector of minimal sample size (so that \( 0 < m_2^N(x) < \infty \), for all \( x \)). It can be verified that the CIP \( p_2^1(\theta_2|\theta_1) \) is always proper, while the intrinsic prior \( p_2^1(\theta_2) \) may be improper. This distinction is very important for the subsequent developments of this paper. The methodology that we will implement, named encompassing prior approach, requires the prior to be proper, and for this reason we will use the CIP. Finally, notice that neither CIP nor IP depends on data.

A CIP is tailored to the comparison of model \( M_2 \) relative to \( M_1 \). In particular, \( p_2^1(\theta_2|\theta_1) \) accumulates more mass than \( p_2^N(\theta_2) \) around the parameter subspace which characterizes \( M_1 \). This is a very reasonable property (Jeffreys, 1961); see also Consonni and La Rocca (2008), Mulder et al. (2010), and Consonni, Moreno, and Venturini (2011). A similar property is of course enjoyed by the intrinsic prior \( p_2^1(\theta_2) \), because it is an average (possibly with respect to an improper measure) of conditional intrinsic priors. CIP and IP are an effective way of “linking” the priors under the two models being compared.

The BF of model \( M_2 \) against \( M_1 \) under the CIP is given by

\[
BF_{21}^I(y|\theta_1) = \frac{\int f_2(y|\theta_2) p_2^1(\theta_2|\theta_1) d\theta_2}{\int f_1(y|\theta_1) p_1^N(\theta_1) d\theta_1}.
\]

A similar calculation could be done under the IP but is omitted because it will not be used in this paper. Clearly, since \( \theta_1 \) is unknown, \( BF_{21}^I(y|\theta_1) \) is of no direct use; however, we like to single it out, because it will play a special role in our method.

The above procedure deals with two nested models. For two nonnested models \( M_1 \) and \( M_2 \), difficulties may arise because intrinsic priors need not exist or need not be unique (Dmochowski, 1996). This suggests to reformulate the original problem as a comparison of two nested models. There are basically two approaches here, usually named encompassing from above and from below (Liang, Paulo, Molina, Clyde, & Berger, 2008). In the former a larger model \( M \) is sought that nests both \( M_1 \) and \( M_2 \), see also Berger and Mortera (1999). However, this approach is not satisfactory
when dealing with intrinsic priors, because it usually leads to two distinct priors under $M$ (one for each for each of the two models under comparison). This difficulty was recognized for instance in Moreno (2005) who suggested using a smaller model $M_0$ which is nested in both $M_1$ and $M_2$ and is of lower dimension (encompassing from below). In this way each model has a unique prior, and the comparison within the two pairs $\{M_1, M_0\}$ and $\{M_2, M_0\}$ can be carried out through $BF_{10}^{M}(y)$ and $BF_{20}^{M}(y)$, from which $BF_{1}^{M}(y) = BF_{2}^{M}(y) \times BF_{10}^{M}(y)$ can be coherently deduced, because $m_0(y)$, the marginal distribution of the data under $M_0$, is the same under the two distinct BF.

2.2. Encompassing and Truncated Priors

Consider a model $M_c$, and let $\theta \in \Theta$ be its parameter. We assume that $\Theta$ is an unrestricted Euclidean space of the appropriate dimension. Define a collection of constrained models $\{M_c\}$. Let $\Theta_c \subset \Theta$ denote the constrained parameter subset characterizing $M_c$. Since $\Theta$ contains each $\Theta_c$, we refer to $M_c$ as the encompassing model.

A natural way to compare the models $M_c$ is to assign a unique proper prior to $\theta$ under $M_c$, $p(\theta|M_c)$, having support $\Theta$. Next, assuming for the moment only inequality constraints, the parameter prior under $M_c$, $p(\theta|M_c)$, can be derived by truncating $p(\theta|M_c)$ to the subspace $\Theta_c \subset \Theta$. Since $M_c$ is defined only through inequality constraints, $\dim(\Theta_c) = \dim(\Theta)$; accordingly, we still denote with $\theta$ the parameter for model $M_c$, and append the model symbol as a conditioning event in the prior. This top-down assignment across parameter spaces, also called encompassing prior approach, establishes a natural link between all priors.

Consider now the BF of model $M_c$, equipped with its restriction prior $p(\theta|M_c)$, versus the encompassing model $M_c$, with prior $p(\theta|M_c)$. It can be checked that

$$BF_{ce}(y) = \frac{Pr(\theta \in \Theta_c|y, M_c)}{Pr(\theta \in \Theta_c|M_c)};$$

see also Klugkist and Hoijtink (2007). The quantity in (3) is the relative belief ratio of subset $\Theta_c$, as described in Baskurt and Evans (2013), and is related to the Savage–Dickey density ratio; see Wetzels, Grasman, and Wagenmakers (2010). Notice the simplicity of this calculation, and how it automatically adjusts for model complexity. In particular, if $\Theta_c$ is very "small" relative to $\Theta$, then both the numerator and denominator of (3) are also likely to be very small; yet $BF_{ce}(y)$ can be substantial, with an upper bound equal to $1/Pr(\theta \in \Theta_c|M_c)$.

The encompassing/truncation approach was presented assuming that the various submodels had been specified exclusively by inequality constraints. The reason is that strict positivity of the numerator and denominator of (3) breaks down if $\Theta_c$ is specified also by means of equality constraints, under standard continuous priors $p(\theta|M_c)$. To solve this difficulty, Klugkist and Hoijtink (2007) advocate the use of “about equality” constraints. This is equivalent to approximating a point hypothesis $\theta = \theta_0$ through an interval hypothesis $|\theta - \theta_0| < b$. Besides being ad hoc, this method raises the usual question of how to fix $b$. Wesel et al. (2011) develop a method to compute the BF of an equality constrained model against the encompassing model through a sequence of “about equality constrained models” corresponding to a decreasing sequence $\{b_r\}, r = 1, \ldots, R$, $R \rightarrow \infty$, until stabilization in the result takes place.

2.3. Bayes Factors and Posterior Model Probabilities

In this subsection we present a novel proposal for constructing objective priors for comparing constrained models, where the constraints can involve inequalities, as well as equalities, with the latter being treated exactly, i.e., without approximations.

Consider a general constrained model $M_c$, possibly involving both equality and inequality constraints, and characterized by a parameter subspace $\Theta_c \subset \Theta$. Define the encompassing-$M_c$
model, written $M_{c(c)}$, as the model whose parameter space $\Theta_{c(c)}$ has the same equality constraints as $\Theta_c$, whereas the inequality constraints are entirely relaxed. This step is named “completion” in Pericchi et al. (2008). Notice that $\dim(\Theta_{c(c)}) = \dim(\Theta_c)$. In particular $M_{c(c)}$ may coincide with a model in the list of entertained models, or it may be a new, additional model; Sect. 3 illustrates this point.

We also introduce the null sampling model $M_0 : \{ f(y|\theta_0, M_0), \theta_0 \in \Theta_0 \}$, with the requirement that it be nested in all the encompassing-$M_c$ models under consideration. We now turn to the specifications of prior distributions.

- For each model $M_c$, identify the corresponding encompassing-$M_c$ model $M_{c(c)}$. Let $p^I(\theta_{c(c)}|M_{c(c)})$ be its default prior.
- Compute the conditional intrinsic prior for $\theta_{c(c)}$, given $\theta_0$, under model $M_{c(c)}$

$$p^I(\theta_{c(c)}|\theta_0, M_{c(c)}),$$

as in (1). Recall that $p^I(\theta_{c(c)}|\theta_0, M_{c(c)})$ is a proper distribution.
- Define the parameter prior under $M_c$, conditional on $\theta_0$, by restricting (4) to the subspace $\Theta_c$

$$p(\theta_{c(c)}|\theta_0, M_c) = k_c(\theta_0)p^I(\theta_{c(c)}|\theta_0, M_{c(c)})1(\Theta_c),$$

where $1/k_c(\theta_0) = \int_{\Theta_c} p^I(\theta_{c(c)}|\theta_0, M_{c(c)})d\theta_{c(c)}$. Recall that $0 < k_c(\theta_0) < \infty$ because $p^I(\theta_{c(c)}|\theta_0)$ is proper.

We are now ready to compute the BF for $M_c$, relative to $M_0$ for every submodel $M_c$.
- Using the encompassing prior approach, we first compute

$$BF_{c,c(c)}^{1}(y|\theta_0) = \frac{Pr^I(\theta \in \Theta_c|y, \theta_0, M_{c(c)})}{Pr^I(\theta \in \Theta_c|\theta_0, M_{c(c)})},$$

as in Equation (3).
- Using the standard intrinsic prior approach for nested models, we compute the BF based on the conditional intrinsic prior $BF_{c(c),0}^{1}(y|\theta_0)$. This can be done as in (2) replacing $\theta_2$ with $\theta_{c(c)}$, and $\theta_1$ with $\theta_0$.
- Finally,

$$BF_{c,0}^{1}(y|\theta_0) = BF_{c,c(c)}^{1}(y|\theta_0) \times BF_{c(c),0}^{1}(y|\theta_0).$$

Notice that $BF_{c,0}^{1}(y|\theta_0)$ is well defined, because the prior for $\theta_{c(c)}$ under model $M_{c(c)}$ is the same in the two BFs appearing on the right-hand side, namely $p^I(\theta_{c(c)}|\theta_0)$, and the same applies to the marginal densities for $y$ under $M_{c(c)}$, which therefore cancel out.

Having obtained $BF_{c,0}^{1}(y|\theta_0)$ as in (7) the posterior probability of $M_c$, given $\theta_0$, is readily available via the formula

$$Pr^I(M_c|y, \theta_0) = \left(1 + \sum_{l \neq c} \frac{p_l}{p_c} BF_{c,0}^{1}(y|\theta_0)\right)^{-1},$$

as in (8).
where \( BF_{lc}^I(y|\theta_0) = BF_{00}^I(y|\theta_0) \times BF_{0c}^I(y|\theta_0) \) with \( BF_{00}^I(y|\theta_0) \) calculated as in (7), and \( p_c = \Pr(M_c) \) is the prior probability of model \( M_c \).

All the calculations performed so far are conditionally on \( \theta_0 \). Eventually, in order to implement our procedure, a value for \( \theta_0 \) has to be fixed. Note that the dimension of \( \theta_0 \) is typically very low, because it indexes the null model. For instance, in the normal ANOVA model considered in the next section \( \theta_0 \) is a two-dimensional vector. Accordingly, one can perform inference on \( \theta_0 \) under \( M_0 \), using the noninformative prior \( p_N(\theta_0|M_0) \), and estimate \( \theta_0 \) through its posterior expectation. With moderate to large sample sizes, the above result will in general be very close to the maximum likelihood estimate \( \hat{\theta}_0 \).

3. ANOVA Models

In this section we provide a detailed analysis of constrained one-way normal ANOVA models using the methodology described in Sect. 2.3. The encompassing sampling model for the observations, conditional on \( (\mu_1, \ldots, \mu_J) \) and \( \sigma \), is

\[
y_{ij} = \mu_j + \epsilon_{ij},
\]

where \( j = 1, \ldots, J \) denotes groups and \( i = 1, \ldots, n_j \) denotes units within group \( j \). Under the encompassing model \( M_e \), the mean structure is unconstrained so that \( (\mu_1, \ldots, \mu_J) \in \mathbb{R}^J \), while the error term satisfies the usual assumption of linear regression models, namely \( \epsilon_{ij} \mid \sigma \sim \text{iid} N(0, \sigma^2) \). For concreteness and motivation, we start by considering an ANOVA model choice setting presented in Lucas (2003), and further analyzed in Wesel et al. (2011); in this way we can illustrate the implementation of our method directly on this problem. The data originate from a psychological experiment measuring the attitude of subjects classified in \( J = 5 \) groups; see also Sect. 4.2. Four models of interest (theories) are identified in terms of relationships among the group means

\[
M_e : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5
\]
\[
M_a : \mu_2 < \mu_1 < \mu_4 \prec \{\mu_3 = \mu_5\}
\]
\[
M_b : \{\mu_1, \mu_3\} > \{\mu_2, \mu_4, \mu_5\}
\]
\[
M_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5,
\]

where \( M_e \) is the encompassing model, \( M_0 \) the null model, and \( M_a \) and \( M_b \) are intermediate models. Let \( y^T = (y_{11}, \ldots, y_{n_1,1}, \ldots, y_{1J}, \ldots, y_{n_J,j}) \) denote the vector of responses. Under the usual normal setup, we can rewrite model \( M_0 \) as

\[
M_0 : y = \alpha_0 1_n + \epsilon_0,
\]

where \( n = (n_1 + \cdots + n_J) \), \( 1_n \) is an \( n \)-dimensional vector with all components equal to 1, \( \alpha_0 \) is the common mean, and \( \epsilon_0 \mid \sigma_0 \sim N_n(0_n, \sigma_0^2 I_n) \), where \( 0_n \) is the \( n \)-dimensional vector with all components equal to 0, and \( I_n \) is the identity matrix of order \( n \). On the other hand, \( M_e \) can be written as

\[
M_e : y = \alpha 1_n + X\delta + \epsilon,
\]
where $\alpha$ is the mean of group 1, $\delta^T = (\delta_2, \ldots, \delta_J)$ represents the additional mean effect of group $j = 2, \ldots, J$, relative to group 1. $X$ is a $n \times (J - 1)$ matrix, with column $j$ containing a one in positions corresponding to units in group $j$ ($j = 2, \ldots, J$), and zero otherwise. Finally, $\epsilon \sim N_n(0_n, \sigma^2 I_n)$.

The mean structure of the constrained model $M_a$ can be rewritten as

$$M_a : \delta_2 < 0, \delta_3 > \delta_4 > 0,$$

where $\delta_3$ is the additive term that appears for units in group $j = 3$ and $j = 5$. We used the convention that, whenever an equality constraint is established among a subset of group means, the corresponding $\delta$ is indexed by the lowest index of the original constituent groups.

On the other hand, the mean structure of the constrained model $M_b$ can be rewritten as

$$M_b : \{\delta_2, \delta_4, \delta_5\} < 0, \delta_3 > \{\delta_2, \delta_4, \delta_5\}.$$

Finally, the encompassing-$M_a$ model, $M_{e(a)}$, can be written as

$$M_{e(a)} : y = \alpha_{e(a)} 1_n + X_{e(a)} \delta_{e(a)} + \epsilon_{e(a)},$$

where $\delta_{e(a)}^T = (\delta_{e(a),2}, \delta_{e(a),3}, \delta_{e(a),4})$ is a three-dimensional vector whose components represent, in the order, the mean excess (relative to group 1) of group 2, of groups 3 and 5, and of group 4. We emphasize that $M_{e(a)}$ is an encompassing model because its parameters are free to vary without constraints.

We now return to the general formulation. Consider a constrained ANOVA model $M_c$. Let $M_{e(c)}$ be the corresponding encompassing-$M_c$ model, and denote its parameter space by $\Theta_{e(c)}$. Let $\Theta_c \subset \Theta_{e(c)}$ be the parameter space of $M_c$ characterizing its inequality constraints relative to $M_{e(c)}$ (recall that $M_c$ is distinguishable from $M_{e(c)}$ only by means of inequality constraints). The goal is to compute the BF of each $M_c$ against the null model $M_0$ based on the conditional intrinsic prior procedure described in Sect. 2.3. This will be achieved in three steps:

1. compute $BF_{M_{e(c)},M_0}^1(y|\alpha_0, \sigma_0)$;
2. compute $BF_{M_c,M_{e(c)}}^1(y|\alpha_0, \sigma_0) = \frac{p_{\{\theta \in \Theta_c|y,\alpha_0,\sigma_0,M_{e(c)}\}}}{p_{\{\theta \in \Theta_c|y,\alpha_0,\sigma_0,M_{c}\}}}$;
3. finally,

$$BF_{M_c,M_0}^1(y|\alpha_0, \sigma_0) = BF_{M_{e(c)},M_0}^1(y|\alpha_0, \sigma_0) \times BF_{M_c,M_{e(c)}}^1(y|\alpha_0, \sigma_0). \tag{10}$$

We will examine the first two steps separately below.

### 3.1. Bayes Factor of an Encompassing Model Relative to the Null Model

Consider a constrained ANOVA model $M_c$ and its encompassing-$M_c$ model $M_{e(c)}$. With slight abuse of terminology, we name the latter $M$ to simplify notation. Clearly $M$ contains only unconstrained parameters, and we write its Bayesian version as

$$\{f(y|\alpha, \delta, \sigma, M) = N_n(y|\alpha 1_n + X\delta, \sigma^2 I_n), \ p^N(\alpha, \delta, \sigma|M) \propto 1/\sigma\}. \tag{11}$$
Consider the comparison of the pair \((M, M_0)\) and the corresponding intrinsic priors. We have

\[
p'(\alpha, \delta, \sigma | \alpha_0, \sigma_0, M) = \frac{2}{\pi \sigma_0 (1 + \frac{\sigma^2}{\sigma_0^2})} N_f(\alpha, \delta | \alpha_0 e, (\sigma^2 + \sigma_0^2)W^{-1}),
\]

\((12)\)

\(\alpha \in \mathbb{R}, \delta \in \mathbb{R}^{J-1}, \sigma > 0\), where \(e^T = (1, 0, \ldots, 0), W^{-1} = \frac{n}{J+1} (Z^T Z)^{-1}\), with \(Z \equiv (1_n; X)\).

Result (12) is standard in the intrinsic prior methodology for normal linear regression models; see, for instance, Girón et al. (2006, formula (4)), and references therein. Recall that the conditional intrinsic prior (12) is proper. Moreover,

\[
(\alpha, \delta | \sigma, \alpha_0, \sigma_0, M) \sim N_f(\alpha_0 e, (\sigma^2 + \sigma_0^2)W^{-1})
\]

\((13)\)

\[
(\sigma^2 | \sigma_0, M) \sim InvBeta\left(\frac{1}{2}, \frac{1}{2}, \sigma_0^2\right),
\]

\((14)\)

where \(InvBeta(a, b, c), a > 0, b > 0, c > 0\), is an inverted-beta density with parameters \((a, b, c)\) having density

\[
p(v | a, b, c) = \frac{c^b}{B(a, b)} v^{a-1} \left(\frac{1}{v + c}\right)^{a+b}, \quad v > 0
\]

(Raiffa & Schlaifer, 1961, p. 221).

Consider now

\[
m^I(y | \alpha_0, \sigma_0, M)
\]

\[
= \int_0^{\infty} \left( \int_{\mathbb{R}^j} f(y | \gamma, \sigma, M) p^I(y | \gamma, \alpha_0, \sigma_0, M) d\gamma \right) p^I(\sigma | \alpha_0, \sigma_0, M) d\sigma
\]

\[
= \int_0^{\infty} \left( \int_{\mathbb{R}^j} N_n(y | Z \gamma, \sigma^2 I_n) N_f(\gamma | \alpha_0 e, (\sigma^2 + \sigma_0^2)W^{-1}) d\gamma \right) p^I(\sigma | \alpha_0, \sigma_0, M) d\sigma,
\]

\((15)\)

where \(Z \equiv (1_n; X)\) and \(\gamma^T \equiv (\alpha, \delta^T)\).

The inner integral in (15) yields \(N_n(y | \alpha_0 Ze, \sigma^2 I_n + (\sigma^2 + \sigma_0^2)ZW^{-1}Z^T)\). This result can be shown directly or applying Lemma 3 of Moreno, Girón, and Torres (2003).

Noticing that \(Ze = 1_n\), we now have to compute

\[
m^I(y | \alpha_0, \sigma_0, M) = \int_0^{\infty} N_n(y | \alpha_0 1_n, \sigma^2 I_n + (\sigma^2 + \sigma_0^2)ZW^{-1}Z^T) \frac{2}{\pi \sigma_0 (1 + \frac{\sigma^2}{\sigma_0^2})} d\sigma.
\]

\((16)\)

An alternative expression for (16) is provided in Appendix, together with an approximate evaluation \(\hat{m}^I(y | \alpha_0, \sigma_0, M)\).
Finally, the approximate BF based on the conditional intrinsic prior is computed as

\[
\hat{BF}_{M_{e(c)}, M_0}^I (y | \alpha_0, \sigma_0) = \frac{\hat{m}^I (y | \alpha_0, \sigma_0, M_{e(c)})}{f (y | \alpha_0, \sigma_0, M_0)} \approx \frac{\hat{m}^I (y | \alpha_0, \sigma_0, M_{e(c)})}{N_h (y | \alpha_0, \sigma_0^2 I_n)}.
\]

The evaluation of (15) requires the full set of individual observations \( y \). Occasionally, only the sufficient statistics are available, namely the group sample means and variances. It is possible to provide an expression for \( m^I (y | \alpha_0, \sigma_0, M) \) based only on the sufficient statistics. Details are provided in Appendix.

### 3.2. Bayes Factor of a Constrained Model Relative to its Encompassing Model

In this section we deal with the computation of

\[
BF_{M_c, M_{e(c)}}^I (y | \alpha_0, \sigma_0) = \frac{Pr^I \{ \theta \in \Theta_c | y, \alpha_0, \sigma_0, M_{e(c)} \}}{Pr^I \{ \theta \in \Theta_c | \alpha_0, \sigma_0, M_{e(c)} \}}. \tag{17}
\]

Consider first the denominator of (17). An analytical evaluation is impossible; however, the conditional intrinsic prior as described in (13) and (14) (with \( M = M_{e(c)} \)) lends itself to an immediate estimate of the denominator, by iteratively sampling values of \( \sigma^2 \) from (14) and then sampling values of \( (\alpha, \delta) \) from the conditional distribution (13). Let \( (\alpha^{(t)}, \delta^{(t)}), t = 1, \ldots, T \) be the sampled values. Estimate the denominator of (17) as

\[
\hat{Pr}^I \{ \theta \in \Theta_c | \alpha_0, \sigma_0, M_{e(c)} \} = \frac{\# \{ \delta^{(t)} \in \Theta_c \}}{T}.
\]

Consider now the numerator of (17). This involves the conditional intrinsic posterior distribution. Letting \( \gamma_{e(c)}^T = (\alpha_{e(c)}, \delta_{e(c)}) \), we have

\[
p^I \left( \gamma_{e(c)}, \sigma_{e(c)}^2 | y, \alpha_0, \sigma_0, M_{e(c)} \right) \propto f (y | \gamma_{e(c)}, \sigma_{e(c)}, M_{e(c)}) \ p^I \left( \gamma_{e(c)} | \alpha_0, \sigma_0, M_{e(c)} \right) \ p^I \left( \sigma_{e(c)}^2 | \alpha_0, \sigma_0, M_{e(c)} \right) \propto N_h (y | \gamma_{e(c)}^T, \sigma_{e(c)}^2 I_n) \ \text{InvBeta} \left( \frac{1}{2}, \sigma_{e(c)}^2 \right).
\]

Since the prior is not conjugate to the likelihood, the posterior is not amenable to iterative direct sampling as for the prior (13) and (14). However, we can resort to an MCMC implementation; see Appendix.

A Reviewer pointed out that the evaluation of (17) based on the number of draws falling inside \( \Theta_c \) may not be precise, especially in high dimensions and for parameter sets having very small probabilities. In this case a more accurate procedure would use MCMC samples coupled with conditional quantities which can be expressed in analytical form; see Morey, Rouder, Pratte, and Speckman (2011) and Mulder (2014b). We have employed this methodology to double check our results in the examples contained in the next section, and found an excellent agreement between the two methods. However, since the MCMC method based on conditional quantities has proved to perform extremely well in a variety of situations, we cannot but recommend its use at least in settings characterized by high dimensions and very low subset probabilities. For completeness, we close this section by adding that the evaluation of the numerator and denominator of (17) may be also tackled using an importance sampling strategy (Bartolucci, Scaccia, & Farcomeni, 2012), although we did not experiment with it.
4. Applications

4.1. Simulation Examples

In this subsection we evaluate the performance of our method through some simulation studies. For comparison purposes we used the same setting presented in Wesel et al. (2011), enriched with an extra simulation scenario to better evaluate the performance of our method. The first example concerns one-way ANOVA experiments for six populations having distinct structures of group means, homogeneous variances (homoscedasticity), and equal group sample sizes. In the second example, group variances are allowed to be different (heteroscedasticity). In the third example, sample sizes are allowed to vary both for the homoscedastic and the heteroscedastic case. Computing time for the examples in this subsection was 20 s for each run on a laptop using a 1.73 GHz processor. The relevant code is available in Online Supplementary Material.

Example 1. We consider six populations \( \{1, 2s, 2m, 2l, 3, 4\} \), each having five groups \( j = 1, \ldots, 5 \). To each population, there is a corresponding true generating model, according to the scheme described in Table 1. We note that each of the populations \( \{2s, 2m, 2l\} \) corresponds to model \( M_2 \) with an increasing separation between adjacent means, so that \( 2s \) is the population with the smallest separation and \( 2l \) that with the largest one.

The competing models are

\[
M_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \\
M_e : \mu_1, \mu_2, \mu_3, \mu_4, \mu_5 \\
M_2 : \mu_1 < \mu_2 < \mu_3 < \mu_4 < \mu_5 \\
M_3 : \mu_2 < \mu_1 < \mu_4 < \{\mu_3 = \mu_5\}.
\]

To validate our method we simulated 500 datasets for each of the six populations \( \{1, 2s, 2m, 2l, 3, 4\} \), and for each of the five groups \( j = 1, \ldots, 5 \), separately for two group sizes \( n_j = 25 \) and \( n_j = 50 \). The choice of 500 simulations was made for comparison purposes with Wesel et al. (2011). We also experimented with 1000, and 2000 simulated datasets, both in this example and in the following Examples 2 and 3, and found essentially the same results. Accordingly, the findings we present in the tables below are based on 500 simulations.

For each model we report the percentage of times (rounded to an integer value) in which the model obtained the highest Bayes factor (10) against the null model \( M_0 \); equivalently the percentage of times it scored the highest posterior probability, because we assume equal prior model probabilities. Values in boldface correspond to the true model. Values in brackets are those
Example 1: percentage over 500 simulations of largest Bayes factors, and posterior model probability medians ($PMP_{med}$) for the correct model.

| Pop | $n_j$ | $M_0$ | $M_2$ | $M_3$ | $M_4$ | $PMP_{med}$ |
|-----|------|------|------|------|------|-------------|
| 1   | 25   | 100  | 0    | 0    | 0    | 1.00        |
|     | 100  | (89) | (2)  | (4)  | (5)  | (0.80)      |
|     | 100  | (96) | (1)  | (2)  | (1)  | (0.93)      |
| 2s  | 25   | 82   | 7    | 5    | 6    | 0.02        |
|     | (6)  | (79) | (13) | (2)  |      | (0.75)      |
|     | (1)  | (92) | (7)  | (0)  | (0.93) |
| 2m  | 25   | 10   | 79   | 4    | 6    | 0.85        |
|     | (0)  | (97) | (3)  | (0)  | (0.94) |
|     | (0)  | (100)| (0)  | (0)  | (0.98) |
| 2l  | 25   | 10   | 80   | 3    | 7    | 0.93        |
|     | (0)  | (99) | (1)  | (0)  | (0.98) |
|     | (0)  | (100)| (0)  | (0)  | (0.99) |
| 3   | 25   | 4    | 0    | 96   | 0    | 0.93        |
|     | (0)  | (0)  | (96) | (4)  | (0.94) |
| 4   | 25   | 0    | 0    | 0    | 100  | 1.00        |
| 50  | 0    | 0    | 0    |      | 0    | 1.00        |

In brackets the corresponding results of Wesel et al. (2011) who did not consider Population 4.

In Wesel et al. (2011, Tables 5 and 6) using an empirical Expected Posterior Prior (EEPP) with an optimal minimal training sample size equal to 2. (Notice that in the latter work the null model is labeled as $M_1$, while the encompassing model is indicated as $M_0$.) The last column reports the median of the posterior model probabilities (out of the 500 simulations) for the correct model ($PMP_{med}$).

Some broad features emerge from Table 2. When the null model $M_0$ holds, our method and EEPP are able to identify it very accurately; a similar conclusion holds when the generating model is $M_3$, and to a good extent also for model $M_2$ under populations $2s$ and $2m$, with our method performing somewhat less satisfactorily than EEPP, especially when $n_j = 25$. Results instead differ for population $2s$ (smallest effect size). Our method favors $M_0$ when $n_j = 25$, while it gives a 50% chance to either $M_0$ or $M_2$ when $n_j = 50$. Notice that according to Cohen’s scale population, $2s$ would be classified as being of medium effect size; the standard $p$-value for testing $M_0$, based on simulations from $2s$ with sample sizes $n_j = 25$, is slightly above the 1% threshold (based on an average of the $p$-values obtained from a small set of runs we conducted), suggesting evidence against the null. The fact that our method favors the null model in this case reveals a discrepancy between the $p$-value and the Bayes factor. This is not new, since behaviors of this kind have been documented in the literature over a long period of time; an early reference being Sellke, Bayarri, and Berger (2001). A more recent reference on this issue is Wagenmakers (2007) who argues that the $p$-value typically overstates the evidence against the null model when compared with the Bayes factor. This point is reinforced in Johnson (2013) who recommends much more stringent thresholds for the $p$-values in order to declare significance of findings.
Example 2. In this example data were generated from the same six populations discussed in the previous example. However, to evaluate sensitivity to model variances, each experiment was replicated under three distinct heteroscedastic settings, characterized by an increasing value of the ratio $F$ between the largest and smallest group variance; see Table 3.

The results are summarized in Table 4 according to the same format of Table 2. The broad conclusion is that our method is still capable of identifying the true generating model, with performances similar to those reported by Wesel et al. (2011) (Tables 8 and 9). This suggests that our model selection procedure is effective also under heteroscedasticity.

Example 3. To further evaluate the robustness of our method, we finally considered heteroscedasticity jointly with unequal group sizes. Specifically for each population, and each of the three values of the ratio $F$, we considered two unequal group size settings: i) increasing ($n_1 = 5$, $n_2 = 15$, $n_3 = 25$, $n_4 = 35$, $n_5 = 45$); and ii) decreasing ($n_1 = 45$, $n_2 = 35$, $n_3 = 25$, $n_4 = 15$, $n_5 = 5$). As shown in Table 5 results are very satisfactory, as they essentially reproduce those which hold under the settings of equal group sizes. This suggests that our methodology is robust to heteroscedasticity even when group sizes differ by a factor of 9 from the largest to the smallest.

4.2. Lucas’ Data

Our third example deals with real data from a psychological experiment. The objective of this study is to find out what group members think about the competence of their leader; see Lucas (2003) for further details. It consists of five groups, each having the same size ($n_j = 30$) but different variances. The five groups are (1) randomly assigned male leader; (2) randomly assigned female leader; (3) male leader assigned on ability; (4) female leader assigned on ability; and (5) institutionalized female leader. The following four models represent substantive research interests:

\[
\begin{align*}
M_a : & \mu_2 < \mu_1 < \mu_4 < \{\mu_3 = \mu_5\} \\
M_b : & \{\mu_1, \mu_3\} > \{\mu_2, \mu_4, \mu_5\} \\
M_0 : & \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 \\
M_e : & \mu_1, \mu_2, \mu_3, \mu_4, \mu_5,
\end{align*}
\]

with $M_0$ and $M_e$ denoting the null, respectively encompassing, model. The data are available only in terms of the sufficient statistics (group sample means and variances). Accordingly, we used the corresponding formula for the marginal likelihood reported in Appendix. The results of our
Table 4.
Example 2: percentage over 500 simulations of largest Bayes factors, and posterior model probability medians (PMPmed) for the correct model.

| Pop | F         | n_j = 25 |         |         | PMPmed | n_j = 50 |         |         | PMPmed |
|-----|-----------|----------|---------|---------|--------|----------|---------|---------|--------|
|     |           | M0 | M2 | M3 | Me |     | M0 | M2 | M3 | Me |
| 1   | 1         | 100 | 0  | 0  | 0  | 1.00 | 100 | 0  | 0  | 0  |
|     |           | (92.4) | (2.6) | (5) | (0.88) | (97) | (1.4) | (1.6) | (0.96) |
| 11  | 100       | 0  | 0  | 0  | 0  | 1.00 | 100 | 0  | 0  | 0  |
|     |           | (89.8) | (4) | (6.2) | (0.89) | (93.4) | (2.6) | (4) | (0.96) |
| 25  | 100       | 0  | 0  | 0  | 0  | 1.00 | 100 | 0  | 0  | 0  |
|     |           | (87.6) | (6.2) | (6.2) | (0.88) | (95.6) | (2) | (2.4) | (0.96) |
| 2s  | 1         | 78 | 1  | 1  | 0.077 | 8 | 92 | 0  | 0  | 0.98 |
|     |           | (3.4) | (95.2) | (1.4) | (0.92) | (0) | (99.8) | (0.2) | (0.97) |
| 11  | 72        | 25 | 3  | 0  | 0.193 | 17 | 82 | 1  | 0  | 0.96 |
|     |           | (3.6) | (95) | (1.4) | (0.93) | (0.2) | (99) | (0.8) | (0.97) |
| 25  | 70        | 28 | 0  | 2  | 0.086 | 18 | 82 | 0  | 0  | 0.98 |
|     |           | (7) | (90.6) | (2.4) | (0.92) | (0.4) | (98) | (1.6) | (0.97) |
| 2m  | 1         | 1  | 99 | 0  | 0  | 0.99 | 0 | 100 | 0  | 0  |
|     |           | (0) | (100) | (0) | (0.98) | (0) | (100) | (0) | (0.99) |
| 11  | 8         | 92 | 0  | 0  | 0.98 | 0 | 100 | 0  | 0  |
|     |           | (0.2) | (99.2) | (0.6) | (0.98) | (0) | (99.8) | (0.2) | (0.99) |
| 2l  | 1         | 0  | 100 | 0  | 0  | 1.00 | 0 | 100 | 0  | 0  |
|     |           | (0) | (100) | (0) | (0.99) | (0) | (100) | (0) | (0.99) |
| 11  | 0         | 100 | 0  | 0  | 0.99 | 0 | 100 | 0  | 0  |
|     |           | (0) | (99.6) | (0.4) | (0.98) | (0) | (100) | (0) | (0.99) |
| 25  | 0         | 100 | 0  | 0  | 0.99 | 0 | 100 | 0  | 0  |
|     |           | (0) | (99.8) | (0.2) | (0.98) | (0) | (99.8) | (0.2) | (0.99) |
| 3   | 1         | 0  | 0  | 100 | 0  | 0.99 | 0 | 0  | 100 | 0  |
|     |           | (0) | (99.6) | (0.4) | (0.98) | (0) | (100) | (0) | (0.99) |
| 11  | 0         | 0  | 100 | 0  | 0  | 1.00 | 0 | 0  | 100 | 0  |
|     |           | (0) | (100) | (0) | (0.99) | (0) | (100) | (0) | (0.99) |
| 25  | 0         | 0  | 100 | 0  | 0  | 1.00 | 0 | 0  | 100 | 0  |
|     |           | (0) | (99.8) | (0.2) | (0.98) | (0) | (100) | (0) | (0.99) |
| 4   | 1         | 0  | 0  | 0  | 100 | 1.00 | 0 | 0  | 100 | 1.00 |
|     |           | (0) | (99.8) | (0.2) | (0.98) | (0) | (100) | (0) | (0.99) |
| 11  | 0         | 0  | 0  | 100 | 1.00 | 0 | 0  | 100 | 1.00 |
|     |           | (0) | (99.8) | (0.2) | (0.98) | (0) | (100) | (0) | (0.99) |
| 25  | 0         | 0  | 0  | 100 | 1.00 | 0 | 0  | 100 | 1.00 |

In brackets the corresponding results of Wesel et al. (2011) who did not consider Populations 3 and 4.

The method are contained in Table 6 along with those obtained by Wesel et al. (2011) (notice that their null and encompassing models are labeled differently from ours; also their BF's are against the encompassing model, whereas all our BF's are relative to the null model; of course the resulting posterior model probabilities are unaffected by this change of reference model). The conclusion is that there is an overwhelming evidence in favor of model $M_a$, which actually represents Lucas’ research hypothesis.

5. Discussion

In this paper we have considered the comparison of models defined through inequality and equality constraints from an objective Bayesian perspective. When assessed with respect to alternative objective Bayes methodologies, notably the empirical expected posterior prior of Mulder (2009), Mulder et al. (2010) and Wesel et al. (2011), our method is grounded in the principled
### Table 5.
Example 3: percentage over 500 simulations of largest Bayes factors, and posterior model probability medians ($\text{PMP}_{\text{med}}$) for the correct model with unequal group sample sizes.

| Pop | $F$ | Increasing group sizes | Decreasing group sizes |
|-----|-----|------------------------|------------------------|
|     |     | $M_0$ | $M_2$ | $M_3$ | $M_e$ | $\text{PMP}_{\text{med}}$ | $M_0$ | $M_2$ | $M_3$ | $M_e$ | $\text{PMP}_{\text{med}}$ |
| 1   | 1   | 100   | 0     | 0     | 0     | 1.00 | 96  | 4   | 0   | 0     | 0.99 |
|     |     | (91.8)| (3.8) | (4.4) | (0.82)|     | (86.8)| (6.4)| (6.8) | (0.81)|     |
| 11  | 1   | 100   | 0     | 0     | 0     | 1.00 | 93  | 5   | 0   | 2     | 0.93 |
|     |     | (94.2)| (2.8) | (3)   | (0.88)|     | (70) | (13.6)| (16.4)| (0.73)|     |
| 25  | 1   | 100   | 0     | 0     | 0     | 1.00 | 88  | 7   | 0   | 5     | 0.90 |
|     |     | (94.2)| (3.2) | (2.6) | (0.88)|     | (63.2)| (13.2)| (23.6)| (0.66)|     |
| 2s  | 1   | 77    | 20    | 1     | 2     | 0.075| 80  | 20  | 0   | 0     | 0.08 |
|     |     | (7.2) | (90.6)| (2.2) | (0.88)|     | (5)  | (92.8)| (2.2) | (0.89)|     |
| 11  | 74  | 26    | 0     | 0     | 0.20  | 76  | 23  | 1   | 0     | 0.07 |
|     |     | (15)  | (83.8)| (1.2) | (0.80)|     | (3.6) | (88.8)| (7.6) | (0.90)|     |
| 25  | 78  | 21    | 0     | 1     | 0.085 | 66  | 28  | 1   | 5     | 0.12 |
|     |     | (21.8)| (77.2)| (1)   | (0.74)|     | (4)  | (85.6)| (10.4)| (0.93)|     |
| 2m  | 1   | 4     | 93    | 0     | 3     | 0.92 | 0   | 95  | 0   | 5     | 0.90 |
|     |     | (0.2) | (99.6)| (0.2) | (0.96)|     | (0)  | (99.4)| (0.6) | (0.97)|     |
| 11  | 0   | 100   | 0     | 0     | 1.00  | 3   | 92  | 0   | 5     | 0.96 |
|     |     | (0.4) | (99.6)| (0)   | (0.95)|     | (0)  | (93.6)| (6.4) | (0.96)|     |
| 2l  | 1   | 0     | 99    | 0     | 1     | 1.00 | 0   | 100 | 0   | 0     | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.98)|     | (0)  | (100)| (0)   | (0.98)|     |
| 11  | 0   | 100   | 0     | 0     | 0.99  | 0   | 100 | 0   | 0     | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (98) | (2)   | (0.98)|     |
| 25  | 0   | 100   | 0     | 0     | 0.99  | 0   | 98  | 0   | 2     | 0.99 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (95.2)| (4.8) | (0.98)|     |
| 3   | 1   | 0     | 0     | 100   | 0     | 0.99 | 0   | 100 | 0   | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.98)|     | (0)  | (100)| (0)   | (0.98)|     |
| 11  | 0   | 0     | 100   | 0     | 1.00  | 0   | 100 | 0   | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (98) | (2)   | (0.98)|     |
| 25  | 0   | 0     | 100   | 0     | 1.00  | 0   | 100 | 0   | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (95.2)| (4.8) | (0.98)|     |
| 4   | 1   | 0     | 0     | 0     | 100   | 1.00 | 0   | 100 | 0   | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (98) | (2)   | (0.98)|     |
| 11  | 0   | 0     | 0     | 100   | 1.00  | 0   | 100 | 0   | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (95.2)| (4.8) | (0.98)|     |
| 25  | 0   | 0     | 0     | 100   | 1.00  | 0   | 100 | 0   | 1.00 |
|     |     | (0)   | (100)| (0)   | (0.97)|     | (0)  | (98) | (2)   | (0.98)|     |

In brackets the corresponding results of Wesel et al. (2011) who did not consider model $M_3$ nor Populations 3 and 4.

### Table 6.
Lucas’ data. Bayes factors (BF) for each of the four models against the null model $M_0$, and Posterior Model Probabilities (PMP).

| Model | BF | PMP |
|-------|----|-----|
| $M_a$ | 5487 | 1.0  |
|       | (4969) | (0.98) |
| $M_b$ | 0.00 | 0.00 |
|       | (1.00) | (0.00) |
| $M_0$ | 1.00 | 0.00 |
|       | (1.00) | (0.00) |
| $M_e$ | 0.00 | 0.00 |
|       | (100) | (0.02) |

The corresponding results of Wesel et al. (2011) are represented in brackets.
methodology of intrinsic priors which has been applied with success in a variety of methodological and applied contexts. Additionally, our method is relatively inexpensive from a computational viewpoint, fully automatic, and treats equality constraints exactly, without resorting to approximate representations.

Recently, Mulder (2014c) presented a new method for testing multiple hypotheses with equality and inequality constraints on the parameters of interest. His method suitably modifies the prior inherent in the fractional Bayes factor (O’Hagan, 1995) by centering it on the boundary of the constrained parameter space under investigation. The resulting prior-adjusted default Bayes factor (BF) outperforms the fractional Bayes factor when testing inequality constrained hypotheses. We compared the prior-adjusted default BF (Type II) with the BF produced by our procedure in the simple setting of testing a sharp versus a one-sided hypothesis in a normal model with known variance. We found that the conditional intrinsic prior produced by our method satisfies all the basic properties mentioned in Mulder (2014c, Sect. 4), namely it is based on a minimal (imaginary) experiment, is centered on the boundary $\theta_0$, and is balanced, in the sense of Jeffreys. We emphasize that in our setup these properties are not externally imposed, but rather follow from the general and well-established methodology of intrinsic priors. Additionally, we proved that the prior-adjusted default BF (Type II) and our BF based on the conditional intrinsic prior are essentially the same as soon as the sample size becomes moderately large. We further investigated the relationship between our approach and that of Mulder (2014c) by analyzing a one-way ANOVA presented in Example 2 of that paper. We checked in particular the behavior of the posterior probabilities of the unconstrained and of the constrained hypotheses when evidence is in favor of the latter. The results we obtained confirm that our Bayes factor behaves as the prior-adjusted default BF, if not marginally better; see Supplementary Material for a careful discussion of this example. Finally, we acknowledge that computationally expediency may suggest to use the method proposed by Mulder (2014c) especially in situations where the derivation of the intrinsic prior is complex.

The simulation studies we conducted showed that our method is robust with respect to heteroscedasticity and unequal group sample sizes. Concerning robustness to the sampling model, assumed to be Gaussian throughout the paper, we remark that the marginal distribution of the observations, i.e., the unconditional distribution wherein the parameters have been marginalized out, is no longer Gaussian, but rather a mixture of normal; see formula (16) or equivalently (18). Using an argument presented in Girón et al. (2006, Sect. 2.1), one can show that the latter is a multivariate elliptical distribution with tails heavier than the normal and hence more resistant to potential outliers. We underscore that the latter distribution (and not the sampling Gaussian family) is the distribution actually employed in the Bayes factor to compare alternative models.

In the examples and application, we have used a uniform prior on model space for the sake of simplicity and comparison with results obtained using alternative approaches. Other choices of priors on model space can of course be used in conjunction with our method; see for instance Scott and Berger (2010) in the context of variable selection, or Carvalho and Scott (2009) and Altomare, Consolani, and La Rocca (2013) for graphical model determination.

Although our method is illustrated with regard to one-way ANOVA, the methodology presented in Sect. 2.3 is general and can be in principle applied to the comparison of any set of constrained models. An obvious candidate would be the family of two-way ANOVA models.

As recalled in Introduction, the idea of computing the Bayes factor of a constrained model by splitting the problem into two components: first the computation of the BF of a suitable encompassing (unconstrained) model against a null model, followed by the BF of the constrained model against the encompassing model, had already been advocated in Pericchi et al. (2008) (PLN). Our method differs significantly from their approach because (i) we use (conditionally) intrinsic priors throughout, which are always proper, while PLN only use improper priors; (ii) having a proper encompassing prior on an unconstrained parameter space, we can meaningfully specify a prior on the constrained parameter space by means of a restriction, whereas this step
is unavailable to PLN; and (iii) in PLN the BF of the constrained model against the null model turns out to be the ordinary BF of the encompassing model against the null model multiplied by a correction factor which, being a probability, is always less than one. On the other hand, our corresponding “correction factor” is a ratio of probabilities with an upper bound that can be greater than one and will likely be so especially for highly constrained models well supported by the data. The bottom line is that PNL always deflates their basic BF, while we can actually inflate ours, so that greater evidence can accrue in favor of more parsimonious models supported by the data; see Mulder (2014c) for a similar general argument.

Acknowledgments

We thank the Editor and an Associate Editor for providing useful suggestions. We also thank two Reviewers for a careful reading of our paper. In particular, one Reviewer offered highly perceptive comments prompting further investigations and comparisons with other methods, eventually leading to an improved paper both in content and presentation. This research was supported by Grant D1 (Università Cattolica del Sacro Cuore, 2013–2014).

Appendix

Approximate Evaluation of Integral (16)

We first provide an alternative expression for the integral (16) which is more suitable for numerical evaluation. Make the change of variable

\[ \eta = \frac{\sigma^2}{\sigma^2 + \sigma_0^2}, \]

so that \( \sigma = \sqrt{\frac{\eta \sigma_0^2}{1 - \eta}}, \) with \( 0 \leq \eta \leq 1, \) and one can write the integrand in (16) as a function of \( \eta. \) In particular the variance–covariance matrix of the normal density becomes \( \frac{\sigma_0^2}{1 - \eta}(\eta I_n + ZW^{-1}Z^T). \)

Additionally, since the prior on \( \sigma^2 \) is \( \text{InvBeta}(\sigma^2 | \frac{1}{2}, \frac{1}{2}, \sigma_0^2) \), it can be checked that the induced prior on \( \eta \) is the \( \text{Beta}(1/2, 1/2) \) distribution. As a consequence we can rewrite (16) as

\[
m^L(y | \alpha_0, \sigma_0, M) = \int_0^1 N_n(y | \alpha_0 1_n, \frac{\sigma_0^2}{1 - \eta}(\eta I_n + ZW^{-1}Z^T)) Beta \left( \eta | \frac{1}{2}, \frac{1}{2} \right) d\eta. \tag{18}
\]

Notice that (18) can be regarded as a marginal density of the “sampling density” \( g(y | \eta, \alpha_0, \sigma_0, M) = N_n(y | \alpha_0 1_n, \frac{\sigma_0^2}{1 - \eta}(\eta I_n + ZW^{-1}Z^T)). \) To evaluate (18) we use a method proposed by Chib (1995), and further extended in Chib and Jeliazkov (2001), to approximate a marginal likelihood using an MCMC algorithm.

More generally, for a given model \( M \) with sampling density \( g(y | \eta, M) \) for the data \( y \), and prior \( p(\eta | M) \), the marginal likelihood \( m(y | M) \) can be written as

\[
m(y | M) = \frac{g(y | \eta, M)p(\eta | M)}{p(\eta | y, M)}. \tag{19}
\]
If $\eta^*$ is a high-density point in the support of the posterior, such as the posterior mode or mean, or the maximum likelihood estimate, an approximate expression for the marginal likelihood is given by

$$\ln \hat{m}(y|M) = \ln g(y|\eta^*, M) + \ln p(\eta^*|M) - \ln \hat{p}(\eta^*|y, M),$$

(19)

where $\hat{p}(\eta^*|y, M)$ is an approximation of the value of the posterior $p(\eta^*|y, M)$ at the point $\eta^*$. The method first runs a Metropolis–Hastings (M–H) algorithm with target distribution $p(\eta|y, M)$ for $N$ iterations producing the draws $(\eta(1), \ldots, \eta(N))$. These are used to compute the value of $\eta^*$, so that the first two terms in the r.h.s. of (19) are available. An estimate of the value of the posterior density at the fixed value $\eta^*$ is given by the following formula:

$$\hat{p}(\eta^*|y, M) = \frac{\sum_{i=1}^{N} \alpha(\eta^{(i)}, \eta^*|y)q(\eta^{(i)}, \eta^*|y)}{\sum_{m=1}^{N} \alpha(\eta^*, \eta^{(m)}|y)},$$

(20)

where $\alpha(\eta, \eta'|y)$ is the acceptance probability of the proposed value $\eta'$ and $q(\eta, \eta'|y)$ is the proposal density for the transition from $\eta$ to $\eta'$, possibly depending on $y$. The sequence of the $\eta^{(i)}$ at the numerator of (20) is drawn from the posterior distribution, while that of the $\eta^{(m)}$ at the denominator is drawn from the proposal $q(\eta, \eta'|y)$. The latter is typically inexpensive to generate (we also chose the same number of draws in the numerator and denominator, although in principle they could be different).

When applied to our case, we chose $q(\eta, \eta'|y) = p(\eta' | M)$ so that the acceptance probability simplifies to

$$\alpha(\eta^{(i)}, \eta^*|y) = \min \left\{ 1; \frac{g(y | \eta^*, \alpha_0, \sigma_0, M)}{g(y | \eta^{(i)}, \alpha_0, \sigma_0, M)} \right\}.$$

Occasionally, data for the ANOVA experiment are available only through the sufficient statistics. For the one-way ANOVA this reduces to the group sample means $y_j$, $j = 1, \ldots, J$. Using the notation of Sect. 3.1, let $\tilde{\gamma} = (\tilde{\alpha}, \tilde{\delta}^T)$ be the MLE’s of $\gamma$, i.e., $\tilde{\alpha} = \tilde{\gamma}_1$, and $\tilde{\delta}_j = \tilde{y}_j - \tilde{\gamma}_1$, $j = 2, \ldots, J$.

The key point is that we can write $(y - Zy)^T (y - Zy) = SSE + (y \tilde{\gamma})^T Z^T Z (y \tilde{\gamma})$, where $SSE = (y - Z\tilde{\gamma})^T (y - Z\tilde{\gamma})$ is the residual sum of squares. It can be checked that $SSE = y^T y - 2\tilde{\gamma}^T Z^T y + \tilde{\gamma}^T Z^T Z \tilde{\gamma}$. Moreover, $(Z^T y)^T = (n\tilde{y}, n_2\tilde{y}_2, \ldots, n_J\tilde{y}_J)$, and $y^T y = \sum_J (n_j - 1)s_j^2 + n_J\tilde{\gamma}_j^2$.

As a consequence, one has

$$N_n(y | Zy, \sigma^2 I_n) = \exp\{-\frac{1}{2\sigma^2} SSE|Z^T Z\}^{\frac{1}{2}} N_J(\tilde{\gamma} | \gamma, \sigma^2(z^T Z)^{-1}).$$

Integrating with respect to the conditional intrinsic prior for $\gamma$ (13), one obtains, either by direct calculation or using Lemma 3 in Moreno et al. (2003),

$$\int_{-\infty}^{\infty} N_J(\tilde{\gamma} | \gamma, \sigma^2(z^T Z)^{-1}) N_J(\gamma | \alpha_0 e, (\sigma^2 + \sigma_0^2)W^{-1}) d\gamma$$

$$= N_J(\tilde{\gamma} | \alpha_0 e, \sigma^2(z^T Z)^{-1} + (\sigma^2 + \sigma_0^2)W^{-1}).$$
Using again the transformation \( \eta = \frac{\sigma^2}{\sigma + \sigma_0^2} \), the marginal likelihood for model \( M \) in (16) based on the sufficient statistics can be written as

\[
m^I(y|\alpha_0, \sigma_0, M) = \int_0^\infty \left[ N_f \left( \hat{\gamma} | \alpha_0 e, \sigma_0^2 \left( \eta (Z^T Z)^{-1} + W^{-1} \right) \right) \right] \times \exp \left\{ -\frac{1 - \eta}{2\sigma_0^2 \eta} \left( \frac{1 - \eta}{2\sigma_0^2 \eta} S E \right) \right\} \frac{\sigma_0^2}{\sigma_0^2 + \eta (Z^T Z)^{-1} + W^{-1}} \exp \left\{ -\frac{1 - \eta}{2\sigma_0^2 \eta} \eta S E \right\} d\eta.
\]

To evaluate the integral (21), one can resort again to the above-described method by Chib and Jeliazkov (2001). The only difference is in the way to write the function \( g(y | \eta, \alpha_0, \sigma_0, M) \) which now reads

\[
g(y | \eta, \alpha_0, \sigma_0, M) = N_f \left( \hat{\gamma} | \alpha_0 e, \sigma_0^2 \left( \eta (Z^T Z)^{-1} + W^{-1} \right) \right) \frac{\sigma_0^2}{\sigma_0^2 + \eta (Z^T Z)^{-1} + W^{-1}} \exp \left\{ -\frac{1 - \eta}{2\sigma_0^2 \eta} \eta S E \right\} d\eta.
\]

Finally, the expression of \( f(y | \alpha_0, \sigma_0, M_0) = N_n(y | \alpha_0 1_n, \sigma_0^2 I_n) \) based on the sufficient statistics is

\[
\left( \frac{1}{2\pi\sigma_0^2} \right)^n \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_j \left( y_{ij} - \bar{y}_j \right)^2 \right\} \exp \left\{ -\frac{n}{2\sigma_0^2} \left( \bar{y} - \alpha_0 \right)^2 \right\} ,
\]

where \( S ST = \sum_j \sum_i (y_{ij} - \bar{y}_j)^2 = \sum_j (n_j - 1)s_j^2 + n_j\bar{y}_j^2 - n\bar{y}^2 \) and \( \bar{y} = \sum_j n_j\bar{y}_j / n \).

**Approximate Evaluation of the BF in Equation (17)**

We describe the MCMC procedure to approximate the numerator of Equation (17). The full conditionals of \( (\gamma, \eta) \) are

- **Full conditional of \( \gamma \)**

\[
p^I(\gamma|\sigma, y, \alpha_0, \sigma_0, M) \propto \left( \begin{array}{c} \sum_j \sum_i (y_{ij} - \bar{y}_j)^2 \\ \sum_j (n_j - 1)s_j^2 + n_j\bar{y}_j^2 - n\bar{y}^2 \end{array} \right) \exp \left\{ -\frac{n}{2\sigma_0^2} \left( \bar{y} - \alpha_0 \right)^2 \right\} , \]

where

\[
\Sigma_\gamma(\sigma) = \left( \frac{W}{\sigma^2 + \sigma_0^2} + \frac{Z^T Z}{\sigma^2} \right)^{-1}, \quad \mu_\gamma(\sigma) = \Sigma_\gamma(\sigma) \left( \frac{W}{\sigma^2 + \sigma_0^2} \alpha_0 e + \frac{Z^T y}{\sigma^2} \right).
\]
• Full conditional of $\eta$

First consider the full conditional of $\sigma^2$

\[
p^I(\sigma^2|y, \alpha_0, \sigma_0, M) \propto N_n(y|Z\gamma, \sigma^2 I_n)N_J(\gamma|\alpha_0 e, (\sigma^2 + \sigma_0^2)W^{-1})
\]

\[
\text{InvBeta}(\sigma^2|\frac{1}{2}, \frac{1}{2}, \sigma_0^2)
\]

\[
\propto \left(\frac{1}{\sigma^2}\right) \frac{n+1}{2} \exp\left\{-\frac{C(\gamma)}{2\sigma^2}\right\}\left(\frac{1}{\sigma^2 + \sigma_0^2}\right)^{J+1}
\]

\[
\exp\left\{-\frac{D(\gamma)}{2(\sigma^2 + \sigma_0^2)}\right\},
\]

where

\[
C(\gamma) = (y - Z\gamma)^T(y - Z\gamma), \quad D(\gamma) = (\gamma - \alpha_0 e)^TW(\gamma - \alpha_0 e).
\]

Now make the change of variable

\[
\sigma^2 \mapsto \eta = \frac{\sigma^2}{\sigma^2 + \sigma_0^2},
\]

and taking into account the Jacobian \(\frac{d\sigma^2}{d\eta} = \frac{1}{1-\eta}\), we obtain

\[
p^I(\eta|y, \alpha_0, \sigma_0, M) \propto \left(\frac{1}{\eta}\right)^{\frac{n+1}{2}}(1-\eta)^{\frac{n+1-j}{2}}
\]

\[
\times \exp\left\{-\frac{1}{2\sigma_0^2} \left[ (1-\eta)\left(C(\gamma) + D(\gamma)\right) + \frac{C(\gamma)}{\eta}\right]\right\},
\]

\[
(22)
\]

Since the normalizing constant of \(p^I(\eta|y, \alpha_0, \sigma_0, M)\) is not analytically available, one must resort to a Metropolis step to sample values of $\eta$.

In conclusion, the MCMC procedure to obtain a numerical approximation of the numerator of (17) can be summarized as follows: For $t = 1, \ldots, T$

(i) sample $\eta^{(t)}$ from $p^I(\eta|y^{(t-1)}, \alpha_0, \sigma_0, M_{e(c)})$ and compute $(\sigma^2)^{(t)} = \sigma_0^2 \frac{\eta^{(t)}}{1-\eta^{(t)}}$;

(ii) sample $\gamma^{(t)} = (\alpha^{(t)}, \delta^{(t)})^T$, from $N_J(\gamma|\mu_{\gamma}(\sigma^{(t)}), \Sigma_{\gamma}(\sigma^{(t)}))$;

(iii) compute

\[
\hat{\Pr}^I\{\theta \in \Theta_c|y, \alpha_0, \sigma_0, M_{e(c)}\} = \frac{\#\{\delta^{(t)} \in \Theta_c\}}{T}.
\]

Finally, obtain

\[
\hat{BF}_{M_e,M_{e(c)}}(y|\alpha_0, \sigma_0) = \frac{\hat{\Pr}^I\{\theta \in \Theta_c|y, \alpha_0, \sigma_0, M_{e(c)}\}}{\hat{\Pr}\{\theta \in \Theta_c|\alpha_0, \sigma_0, M_{e(c)}\}}.
\]
When the data are available only in terms of sufficient statistics, the previous algorithm still applies, the only modification being to rewrite the expression for $C(\gamma)$ as $C(\gamma) = SSE + (\hat{\gamma} - \gamma)^T Z^T Z (\gamma - \hat{\gamma})$, and recalling that $(Z^T y)^T = (n \bar{y}, n_2 \bar{y}_2, \ldots, n J \bar{y}_J)$.

References

Altomare, D., Consonni, G., & La Rocca, L. (2013). Objective Bayesian search of Gaussian directed acyclic graphical models for ordered variables with non-local priors. *Biometrics*, 69, 478–487.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions. The theory and application of isotonic regression*. London/New York/Sydney: Wiley.

Bartolucci, F., Scaccia, L., & Farcomeni, A. (2012). Bayesian inference through encompassing priors and importance sampling for a class of marginal models for categorical data. *Computational Statistics & Data Analysis*, 56, 4067–4080.

Baskurt, Z., & Evans, M. (2013). Hypothesis assessment and inequalities for Bayes factors and relative belief ratios. *Bayesian Analysis*, 8, 569–590.

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis, I*, 385–402.

Berger, J. O., & Mortera, J. (1999). Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association*, 94, 542–554.

Berger, J. O., & Pericchi, L. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91, 109–122.

Carvalho, C. M., & Scott, J. G. (2009). Objective Bayesian model selection in Gaussian graphical models. *Biometrika*, 96, 497–512.

Casella, G., Girón, F. J., Martínez, M. L., & Moreno, E. (2009). Consistency of Bayesian procedures for variable selection. *Annals of Statistics*, 37, 1207–1228.

Casella, G., & Moreno, E. (2006). Objective Bayesian variable selection. *Journal of the American Statistical Association*, 101, 157–167.

Casella, G., & Moreno, E. (2009). Assessing robustness of intrinsic tests of independence in two-way contingency tables. *Journal of the American Statistical Association*, 104, 1261–1271.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association*, 96, 270–281.

Consonni, G., Moreno, E., & Venturini, S. (2011). Testing Hardy–Weinberg equilibrium: An objective Bayesian analysis. *Statistics in Medicine*, 30, 62–74.

Consonni, G., & La Rocca, L. (2008). Tests based on intrinsic priors for the equality of two correlated proportions. *Journal of the American Statistical Association*, 103, 1260–1269.

Consonni, G., & Veronese, P. (2008). Compatibility of prior specifications across linear models. *Statistical Science*, 23, 332–353.

Dmochowski, J. (1996). Intrinsic priors via Kullback–Leibler geometry. In *Bayesian statistics, 5 (Alicante, 1994)* (pp. 543–549). New York: Oxford Science Publications/Oxford University Press.

Girón, F. J., Martínez, M. L., Moreno, E., & Torres, F. (2006). Objective testing procedures in linear models: Calibration of the p-values. *Scandinavian Journal of Statistics*, 33, 765–784.

Hojtink, H. (2013). Objective Bayes factors for inequality constrained hypotheses. *International Statistical Review*, 81, 207–229.

Hojtink, H., Klugkist, I., & Boelen, P. A. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford University Press. Corrected impression (1966).

Johnson, V. E. (2013). Revised standards for statistical evidence. *Proceedings of the National Academy of Sciences*, 110, 19313–19317.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773–795.

Klugkist, I., & Hoijtink, H. (2007). The Bayes factor for inequality and about equality constrained models. *Computational Statistics & Data Analysis*, 51, 6367–6379.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychometrical Methods, 10*, 477–493.

Laudy, O., & Hoijtink, H. (2007). Bayesian methods for the analysis of inequality constrained contingency tables. *Statistical Methods in Medical Research*, 16, 123–138.

Leon-Novelo, L., Moreno, E., & Casella, G. (2012). Objective Bayes model selection in probit models. *Statistics in Medicine*, 31, 353–365.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., & Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103, 410–423.

Lucas, J. W. (2003). Status processes and the institutionalization of women as leaders. *American Sociological Review*, 68, 464–480.

Moreno, E. (1997). Bayes factors for intrinsic and fractional priors in nested models. Bayesian robustness. In Y. Dodge (Ed.), *Statistical procedures and related topics* (pp. 257–270). Hayward: Institute of Mathematical Statistics.

Moreno, E. (2005). Objective Bayesian analysis for one-sided testing. *Test*, 14, 181–198.
Moreno, E., Girón, F. J., & Casella, G. (2010). Consistency of objective Bayes factors as the model dimension grows. *Annals of Statistics*, 38, 1937–1952.

Moreno, E., Girón, F., & Torres, F. (2003). Intrinsic priors for hypothesis testing in normal regression models. *Revista de la Real Academia de Ciencias Series A*, 97, 53–61.

Morey, R. D., Rouder, J. N., Pratte, M. S., & Speckman, P. L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology*, 55, 368–378.

Mulder, J. (2014a). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology*, 67, 153–171.

Mulder, J. (2014b). Bayes factors for testing order-constrained hypotheses on correlations. *Journal of Mathematical Psychology*, 72, 104–115.

Mulder, J. (2014c). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. *Computational Statistics and Data Analysis*, 71, 448–463.

Mulder, J. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, 53, 530–546.

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference*, 140, 887–906.

O’Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B (Methodological)*, 57, 99–138.

O’Hagan, A., & Forster, J. (2004). *Kendall’s advanced theory of statistics, Vol. 2b: Bayesian inference* (2nd ed.). London: Arnold.

Pérez, J. M., & Berger, J. O. (2002). Expected posterior prior distributions for model selection. *Biometrika*, 89, 491–512.

Pericchi, L. R. (2005). Model selection and hypothesis testing based on objective probabilities and Bayes factors. In D. Dey & C. Rao (Eds.), *Bayesian thinking modeling and computation, vol. 25 of handbook of statistics* (pp. 115–149). Amsterdam: Elsevier.

Pericchi, L. R., Liu, G., & Núñez Torres, D. (2008). Objective Bayes factors for informed hypotheses: “Completing” and “splitting” the Bayes factors. In H. Hoijtink, I. Klugkist, & P. A. Boelen (Eds.), *Bayesian evaluation of informative hypotheses* (pp. 131–154). New York: Springer.

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. Boston: Clinton Press Inc.

Robertson, T., Wright, F., & Dykstra, R. (1988). *Order restricted statistical inference*. Chichester: Wiley.

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *The Annals of Statistics*, 38, 2587–2619.

Sellke, T., Bayarri, J. M., & Berger, J. O. (2001). Calibration of rho values for testing precise null hypotheses. *The American Statistician*, 55, 62–71.

Silvapulle, M., & Sen, P. (2005). *Constrained statistical inference: Order, inequality, and shape constraints*. London/New York/Sydney: Wiley.

Stern, H. S. (2005). Model inference or model selection: Discussion of Klugkist, Laudy, and Hoijtink (2005). *Psychological Methods*, 10, 494–499.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.

Wesel, F. V., Hoijtink, H., & Klugkist, I. (2011). Choosing priors for constrained analysis of variance: Methods based on training data. *Scandinavian Journal of Statistics*, 38, 666–690.

Wetzels, R., Grasman, R. P., & Wagenmakers, E.-J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio. *Computational Statistics & Data Analysis*, 54, 2094–2102.

*Manuscript Received: 29 APR 2015*

*Final Version Received: 5 MAR 2016*

*Published Online Date: 4 OCT 2016*