This is a pre-publication version of our paper.

The copy-right of the published authoritative version of the paper is owned by:

Article: Bayesian Evaluation of Informative Hypotheses for Multiple Populations Journal: British Journal of Mathematical and Statistical Psychology

Article DOI: 10.1111/bmsp.12145

Bayesian Evaluation of Informative Hypotheses for Multiple Populations

Herbert Hoijtink

Department of Methodology and Statistics, Utrecht University

Xin Gu

Department of Geography and Planning, University of Liverpool

Joris Mulder

Department of Methodology and Statistics, Tilburg University

Author Note

Herbert Hoijtink, Department of Methodology and Statistics, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, The Netherlands. E-mail: H.Hoijtink@uu.nl. The first author is supported by the Consortium on Individual Development (CID) which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.001.003). Xin Gu, Department of Geography and Planning, University of Liverpool, GuXin57@hotmail.com. Joris Mulder, Department of Methodology and Statistics, Tilburg University, J.Mulder3@uvt.nl. The third author is supported by a NWO Vidi Grant (number 452-17-006).

Abstract

The software package Bain can be used for the evaluation of informative hypotheses with respect to the parameters of a wide range of statistical models. For pairs of hypotheses the support in the data is quantified using the approximate adjusted fractional Bayes factor (BF). Currently, the data have to come from one population or have to consist of samples of equal size obtained from multiple populations. If samples of unequal size are obtained from multiple populations, the BF can be shown to be inconsistent. In this paper it will be elaborated how the approach implemented in Bain can be generalized such that multiple population data can properly be processed. The resulting multiple population approximate adjusted fractional Bayes factor (MBF) is implemented in the R package Bain.

Keywords: Bain, Bayes Factor, Informative Hypotheses, Multiple Populations

Bayesian Evaluation of Informative Hypotheses for Multiple Populations

Introduction

This paper is the most recent addition to a sequence of papers in which an alternative for null-hypothesis significance testing has been developed. Important landmarks in this development were: Klugkist, Laudy, and Hoijtink (2005a) and Kuiper, Klugkist, and Hoijtink (2010) who added order constrained hypotheses to the classical null hypothesis and showed in the context of ANOVA models how these can be evaluated using the Bayes factor (Kass and Raftery, 1995); Mulder, Hoijtink, and de Leeuw (2012) who generalized the approach to Bayesian evaluation of informative hypotheses (Hoijtink, 2012), that is, hypotheses specified using equality and inequality (or order) constraints among the parameters of multivariate normal linear models; Gu, Mulder, Dekovic, and Hoijtink (2014) who developed a Bayes factor for the evaluation of inequality constrained hypotheses in a rather wide range of statistical models; and, Mulder (2014) and Gu, Mulder, and Hoijtink (2017), who generalized the latter Bayes factor into the approximate adjusted fractional Bayes factor (AAFBF from now on abbreviated to BF) which can be used to evaluate informative hypotheses for one population data for a wide range of statistical models like normal linear models, logistic regression models, confirmatory factor analyis, and structural equation models.¹

The BF is simple to compute and the only input needed are estimates of the model parameters, the corresponding covariance matrix, and the sample size. However, as will be elaborated in this paper, the BF is inconsistent if samples of unequal size are obtained from multiple populations (similar as O'Hagan's, 1995, fractional Bayes factor as is shown by De Santis and Spezzaferri, 2001). In this paper it will be elaborated how the approximate adjusted fractional Bayes factor, i.e., BF, can be generalized into the multiple population approximate adjusted fractional Bayes factor, i.e., MBF. This Bayes factor is simple to

¹The interested reader is referred to http://informative-hypotheses.sites.uu.nl/ where all the books, dissertations, papers, and software produced during the course of this development are presented.

BAIN FOR MULTIPLE POPULATIONS

compute too, the only input needed are estimates of the model parameters, separate estimates of the corresponding covariance matrix for each population, and the sample size obtained from each population. As will be shown, the MBF is consistent and can therefore be used for testing informative hypotheses with respect to multiple populations.

With the availability of the MBF (and corresponding software) researchers have a viable alternative for null-hypothesis significance testing. In a wide range of statistical models, the null-hypothesis can be replaced by informative hypotheses, and the p-values can be replaced by the MBF. The recent and current critical appraisal of null-hypothesis significance testing in the literature will not be reiterated here. However, the interested reader is referred to Cohen (1994) who called the null-hypothesis the nill hypothesis because he could not come up with research examples in which the null hypothesis might be a realistic representation of the population of interest. This point of view was further elaborated by Royal (1997, pp. 79-81) who claims that the null hypothesis cannot be true, and consequently, that data are not needed in order to be able to reject it. However, the interested reader is also referred to Wainer (1999) who highlights that there are situations where, without dispute, the null hypothesis is relevant. Landmark papers criticizing the use of p-values and significance levels are Ioannides (2005) and Wagenmakers (2007), among others. The latter paper also motivates and illustrates the replacement of p-values by Bayesian hypothesis testing using the BIC (Schwartz, 1987; Raftery, 1995). However, the interested reader is also referred to the American Statistical Association's statement on p-values (Wasserstein and Lazar, 2016) which gives a to the point and balanced overview of what can and can't be done with p-values, and to Benjamin et al. (2018) who propose to redefine statistical significance.

The focus of this paper is on the evaluation of informative hypotheses using Bayes factor. Note that, model selection criteria like AIC and BIC (Schwartz, 1987; Raftery, 1995) cannot be used (Mulder, Klugkist, Meeus, van de Schoot, Selfhout, and Hoijtink, 2009, Section 3). The penalty for model complexity in both criteria is a function of the number of parameters in the model at hand. Since the number of parameters in an unconstrained hypothesis, e.g. $H_u: \theta_1, \theta_2, \theta_3$ is the same as in a constrained hypothesis, e.g., $H_1: \theta_1 > \theta_2 > \theta_3$, it does not reflect that H_1 is more parsimonious than H_u . This problem is solved by Kuiper and Hoijtink (2013) who present the GORIC wich is a generalization of the AIC with a penalty term that does properly reflect that H_1 is more parsimonious than H_u . However, GORIC can only be applied in the context of the multivariate normal linear model, while, as elaborated above, the range of application of the (M)BF is not limited to the multivariate normal linear model.

Also, as is elaborated in Van de Schoot, Hoijtink, Romeijn, and Brugman (2012), the penalty for model complexity used by the DIC (Spiegelhaler, Best, Carlin, and Van der Linde, 2002) is also not suited too quantify how parsimonious an informative hypothesis is. Using a modification of the loss function used by the DIC they obtain the PIC which in the examples provided can be used to evaluate informative hypotheses. However, as was shown by Mulder (2014) using the Bayes factor results in more desirable selection behavior when testing constrained hypotheses than using the PIC.

Silvapulle and Sen (2004) show how, so-called, Type A testing problems (evaluating a null-hypothesis against an informative hypothesis) and Type B testing problems (evaluating an informative hypothesis against an unconstrained hypothesis) can be evaluated using p-values in a wide range of statistical models. Those in favour of null-hypothesis significance testing are well advised to consult this book and the R packages restriktor and ic.infer. The main limitation of this approach is that it can not be used to directly compare two competing informative hypotheses.

Stern (2005) proposes to use the posterior density of H_k for k = 1, ..., K to select the best hypothesis. However, as is elaborated in Klugkist, Laudy, and Hoijtink (2005b), this amounts to using f_k to select the best hypothesis, that is, the complexity c_k is ignored. This will work if each hypothesis has the same complexity. However, if, for example, H_u is compared to H_1 , irrespective of the data, H_u will always be preferred because it has by definition a larger fit than H_1 (cf. Equation 7).

This paper starts with an introduction of the BF and, using a simple two group setup, it will be shown and illustrated that it may show inconsistent behavior if samples of unequal size are obtained from multiple populations. Subsequently, the BF will be generalized into the MBF and, using the same two group setup, it will be shown and illustrated that the MBF does not show inconsistent behavior if samples of unequal size are obtained from multiple populations. Further illustrations of the approach proposed in the context of an analysis of covariance model and a logistic regression analysis will be provided. Illustrations are executed using the R package² Bain³. The R codes and data used in this paper can be found at the bottom of the Bain website (click on the title of this paper). The paper is concluded with a short discussion and contains an Appendix with a further discussion of the consistency of the MBF.

The Approximate Adjusted Fractional Bayes Factor

Consider a model where $\boldsymbol{\theta}$ is a vector of length J containing the structural parameters, and $\boldsymbol{\omega}$ a scalar, vector, or matrix containing the nuisance parameters. Hypotheses can be formalized as:

$$H_k: S_k \boldsymbol{\theta} = s_k, R_k \boldsymbol{\theta} > r_k, \text{ for } k = 1, \dots, K,$$
(1)

where S_k is a $p_k \times J$ matrix imposing p_k equality constraints on $\boldsymbol{\theta}$, R_k is a $q_k \times J$ matrix imposing q_k inequality constraints, and s_k and r_k are vectors containing constants of size p_k and q_k , respectively. Additionally of interest is the unconstrained hypothesis $H_u : \boldsymbol{\theta}$, that is, a hypothesis without constraints on the parameters $\boldsymbol{\theta}$. As will be elaborated below, this hypothesis has a central role in the computation of the Bayes factor.

Mulder (2014), Gu, Mulder, Dekovic, and Hoijtink (2014), Gu (2016), and Gu, Mulder, and Hoijtink (2017) show that the relative support in the data for H_k and H_u can

²https://www.r-project.org/

³https://informative-hypotheses.sites.uu.nl/software/bain/

be quantified using the approximate adjusted fractional Bayes factor:

$$BF_{ku} = \frac{f_k}{c_k},\tag{2}$$

that is the ratio of the fit and the complexity of H_k relative to H_u . The interested reader should consult the references given for the derivation of Equation 2 from the ratio of the marginal likelihoods of H_k and H_u . The Bayes factor from Equation 2 is a quantification of the relative support in the data for H_k against H_u . If, for example, BF_{ku} = 5, the support in the data is five times larger for H_k than for H_u . It will now be elaborated how the BF can be computed and why it is called the approximate adjusted frational Bayes factor. It will also be highlighted that the BF is a member of the family of Bayes factors based on encompassing priors (Klugkist, Kato, and Hoijtink, 2005; Wetzels, Grasman, and Wagenmakers, 2010), that is, Bayes factors for which the prior distribution of the model parameters under H_k is derived from the prior distribution under H_u .

Before providing the formulas for f_k and c_k , it has to be highlighted that the density of the data can be factored according to O'Hagan (1995) as:

$$p(\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{X}) = p(\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{X})^{1-b} p(\boldsymbol{Y} \mid \boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{X})^{b},$$
(3)

where, \mathbf{Y} denotes the data that are modeled (e.g., the dependent variable in a multiple regression) and \mathbf{X} the data that are not modeled and considered to be fixed (e.g., the predictor variables in a multiple regression). The idea of fractional Bayes factors is to use a fraction b of the information in the likelihood function to specify the prior distribution. Usually the fraction b is chosen such that it corresponds to the size of a minimal training sample (Berger and Pericchi, 1996, 2004). For the evaluation of informative hypotheses we implemented in the R package Bain $b = J^*/N$, where J^* denotes the number of independent constraints in $[S_1, R_1, ..., S_K, R_K]$ and N the sample size. This choice can be illustrated using a simple example. If $H_1 : \theta_1 > \theta_2 > \theta_3$ and $H_2 : \theta_1 = \theta_2 = \theta_3$, the number of independent constraints $J^* = 2$, that is, there are two underlying parameters that are combinations of the target parameters with respect to which hypotheses are formulated: $\theta_1 - \theta_2$ and $\theta_2 - \theta_3$. Our choice is motivated by the fact that in the normal linear model, the minimal training sample needed to obtain a proper posterior distribution is equal to the number of parameters. If, for example, a variable is modeled using a normal distribution with unknown mean μ and variance σ^2 , the minimum training sample needed to obtain a proper posterior based on the prior $h(\mu, \sigma^2) = 1/\sigma^2$ is two (c.f., Berger and Pericchi, 2004, Example 1). If, a variable is a linear combination of two predictors with normal error, there are four parameters (intercept, two regression coefficients, residual variance) and, consequently, the minimum training sample equals four.

Gu, Mulder, Dekovic, and Hoijtink (2014), Gu (2016), and Gu, Mulder, and Hoijtink (2017) show that based on Equation 3 and an improper uniform prior for $\boldsymbol{\theta}$, a large sample approximation (see, Gelman, et al., 2013, Chapter 4) of the posterior distribution of $\boldsymbol{\theta}$ under H_u can be obtained:

$$g_u(\boldsymbol{\theta} \mid \boldsymbol{Y}, \boldsymbol{X}) \approx \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}), \tag{4}$$

where $\hat{\theta}$ denotes the maximum likelihood estimate of θ and $\hat{\Sigma}_{\theta}$ the corresponding covariance matrix. Note that, the "approximate" in the name approximate adjusted fractional Bayes factor reflects that for its computation a normal approximation of the posterior distribution is used. An implication of the approximation is that the BF can only be used if a normal approximation to the posterior distribution of θ is reasonable. If the sample size is not too small (see below), this is the case with unbounded parameters like means and regression coefficients as they appear in generalized linear models and structural equation models. This is also the case for the fixed regression coefficients (the random effect would be treated as nuisance parameters) in, for example, two level models. In the latter case, the sample size used is the number of level two units (and not the number of observations of the dependent variable). This is not necessarily the case with naturally bounded parameters like variances (naturally bounded to be larger than zero) and probabilities (naturally bounded between zero and one), although even there, if the sample size is large, a normal approximation of the posterior distribution may be accurate. The interested reader is referred to Gu, Mulder, Dekovic, and Hoijtink (2014) who show that for the evaluation of inequality constrained hypotheses in the context of a multiple regression with two predictors, the difference between the approximate BF implemented in **Bain** and the corresponding non-approximate BF implemented in **Biems** (Mulder, Hoijtink, and de Leeuw, 2012) is negligible if the sample size is at least 20. They also show that inequality constrained hypotheses with respect to the probabilities in a two by two contingency table render an approximate BF that is very similar the the non-approximate BF presented by Klugkist, Laudy, and Hoijtink, (2010), if the sample size is at least 40. Although these results give confidence in the performance of the approximate adjusted fractional Bayes factor, further research in the context of different models is needed in order to fortify these results.

The prior distribution of θ has a covariance matrix which is based on a fraction b of the information in Equation 3 and a mean

$$\boldsymbol{\theta}_B \in \{S_k \boldsymbol{\theta}_B = s_k, R_k \boldsymbol{\theta}_B = r_k\} \text{ for } k = 1, \dots, K,$$
(5)

that is, θ_B denotes a value of θ on the boundary of all the hypotheses under investigation (Mulder, 2014):

$$h_u(\boldsymbol{\theta} \mid [\boldsymbol{Y}, \boldsymbol{X}]^b) = \mathcal{N}(\boldsymbol{\theta}_B, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}/b), \tag{6}$$

where, $[\mathbf{Y}, \mathbf{X}]^b$ stresses that the prior distribution is based on a fraction b of the information in the data. Note that, $\boldsymbol{\theta}_B$ is called the adjusted mean (Mulder, 2014) of the prior distribution, which explains the "adjusted" in the name approximate adjusted fractional Bayes factor. As was shown by Mulder (2014), if, for example, $H_1 : \boldsymbol{\theta} > 0$ is compared with $H_u : \boldsymbol{\theta}$, it holds that the more the data support H_1 the smaller the support in the fractional Bayes factor for H_1 ! This phenomenon is addressed if the adjusted fractional Bayes factor is used, that is, if the prior mean is in agreement with Equation 5, it holds that the more the data are in agreement with H_1 the larger the support in the adjusted fractional Bayes factor for H_1 (see, Mulder, 2014, for further elaborations). Note furthermore, that $h_u(\boldsymbol{\theta} \mid .)$ is a so-called encompassing prior, that is, the prior distribution of $\boldsymbol{\theta}$ under H_k is proportional to $h_u(\boldsymbol{\theta} \mid .)I_{\boldsymbol{\theta} \in H_k}$, where the indicator function is 1 if the argument is true and 0 otherwise (Klugkist, Kato, and Hoijtink, 2005; Wetzels, Grasman, and Wagenmakers, 2010).

There are situations in which there is no solution to Equation 5. For example, if hypotheses are specified using range constraints, e.g., $H_1 : |\theta| < .2$, that is, $H_1 : \theta > -.2, \theta < .2$, there is no solution. **Bain** addresses this problem in the following manner: in Equation 5 (and only in this equation) this (part of an) hypothesis is represented as $H_1 : \theta = 0$, that is, θ_B will be equal to the midpoint of the range specified. The rational is that H_1 essentially implies that $\theta \approx 0$. Another example is given by the hypotheses $H_1 : \theta = 0$ and $H_2 : \theta > 2$. Although each of these hypotheses can be evaluated by itself, they can not be compared using the approximate adjusted fractional Bayes factor because there is no solution to Equation 5, that is, both hypotheses are not compatible because $h_u(.)$ is different for each hypothesis (Hoijtink, 2012, Section 9.9.2.1.). Testing non-compatible hypotheses can be done using BIEMS (Mulder et al., 2012) by instructing the program to use the same unconstrained prior for each of the hypotheses under consideration.

Based on Equations 4 and 6 the relative fit and complexity from Equation 2 are defined as

$$f_{k} = \int_{\boldsymbol{\theta} \in H_{k}} g_{u}(\boldsymbol{\theta} \mid \boldsymbol{Y}, \boldsymbol{X}) d\boldsymbol{\theta} \approx \int_{\boldsymbol{\theta} \in H_{k}} \mathcal{N}(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}) d\boldsymbol{\theta},$$
(7)

and

$$c_{k} = \int_{\boldsymbol{\theta} \in H_{k}} h_{u}(\boldsymbol{\theta} \mid [\boldsymbol{Y}, \boldsymbol{X}]^{b}) d\boldsymbol{\theta} = \int_{\boldsymbol{\theta} \in H_{k}} \mathcal{N}(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{B}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}/b) d\boldsymbol{\theta},$$
(8)

respectively. The interested reader is referred to Gu (2016, Chapter 3) for the algorithms with which the fit and complexity are computed. The strength of the BF lies in its simplicity. Its computation is based only on maximum likelihood estimates and the corresponding asymptotic covariance matrix, and the choice of the fraction b, which is completely determined by the sample size N and the number of independent constraints J^* . The approximate adjusted fractional Bayes factor, also in the paragraphs that follow abbreviated as BF, falls in the category of default, automatic, or pseudo Bayes factors because no priors have to be manually specified. Instead, the prior is automatically constructed using a small fraction of the data, while the remaining fraction is used for hypothesis testing, similar as the fractional Bayes factor (O'Hagan, 1995). The BF is coherent in the sense that $BF_{uk} = 1/BF_{uk}$ and $BF_{kk'} = BF_{ku}/BF_{k'u}$ (O'Hagan, 1997, Sections 3.1 and 3.2). Note that these coherency properties not necessarily hold for other default Bayes factors (O'Hagan, 1997; Robert, 2007, p. 240).

As further noted by Robert (2007, p. 242), a potential issue of the fractional Bayes factor, and therefore also of the BF in Equation 2, is that there is no clear-cut procedure to choose the fraction b. We believe however that the use of a minimal fraction is reasonable as it results in a minimally informative default prior while maximal information in the data is used for hypothesis testing (Berger & Mortera, 1995). Furthermore it has been shown that this choice results in consistent testing behavior (O'Hagan, 1995; Mulder, 2014). Nevertheless, further research about the choice of b would fortify the approach we present in this paper. The interested reader is referred to Gu, Hoijtink, and Mulder (2016), for one evaluation of the choice of b. Another potential issue highlighted by Robert is that default Bayes factors can be computationally intensive (Robert, 2007, p. 242). The BF procedure that is proposed here however is very easy to compute: only the maximum likelihood estimates, error covariance matrix and sample size are needed (Gu et al., 2017).

Finally it is important to note that default Bayes factors may behave as ordinary Bayes factors based on on so-called intrinsic priors (Berger & Pericchi, 1996). Currently however intrinsic priors have not yet been explored for the BF. Although this too is a topic worthy of further research, from a pragmatic point of view it is more important to know whether the BF is consistent, that is, whether the support for the true hypothesis goes to infinity when the sample size grows to infinity. According to O'Hagan (1997, Section 2.1) if hypotheses are nested (in the cases we consider all hypotheses are nested within H_u) and if $b \to 0$ if $N \to \infty$ (which holds for our b), the fractional Bayes factor is consistent. However, De Santis and Spezzaferi (2001) show that the fractional Bayes factor may show inconsistent behavior if data from multiple populations are sampled. Similarly, as shown in the next section, the BF is also inconsistent if the data are sampled from multiple populations. In line with the solution proposed by De Santis and Spezzaferri (2001) for the fractional Bayes factor, the MBF is an extension of the BF that is consistent when testing hypotheses in the case of multiple populations.

Consistency of the Approximate Adjusted Fractional Bayes Factor

When discussing consistency of the (M)BF this will be done in terms of (M)BF_{ku} if H_k is specified using only equality constraints. In this case a Bayes factor is called consistent if (M)BF_{ku} $\rightarrow \infty$ (or 0) when H_k (or H_u) is true and $N_g \rightarrow \infty$ with the same rate for all $g = 1, \ldots, G$. If H_k is specified using only inequality constraints, this will be done in terms of (M)BF_{kc}, where H_c : not H_k , that is the complement of H_k . In this situation, a Bayes factor is called consistent if BF_{kc} $\rightarrow \infty$ (or 0) when H_k (or H_c) is true, as $N_g \rightarrow \infty$ with the same rate for all $g = 1, \ldots, G$. Both scenarios imply that the Gpopulations are treated as one population from which a sample of increasing size (proportionally increasing the sample sizes from each of the G populations) is taken. Note that BF_{kc} = $\frac{BF_{ku}}{BF_{cu}}$, where the numerator and denominator can be computed using Equation 2. Note furthermore, that for hypotheses specified using only equality constraints $H_u = H_c$.

When $BF_{ku} \rightarrow \infty$ or $BF_{kc} \rightarrow \infty$ for the same limit, the Bayes factor is called inconsistent. Another form of inconsistency that will be considered in this paper is whether $(M)BF_{ku} \rightarrow \infty$ or 0, and, $(M)BF_{kc} \rightarrow \infty$ or 0, as $N_g \rightarrow \infty$ for some populations but not all G populations. This situation applies if a sample of increasing size if obtained from some of the G populations whilst the sample size from the other populations remains fixed. De Santis and Spezzaferri (2001) showed for this limit that the fractional Bayes factor (O'Hagan, 1995) is inconsistent. In this section it will be illustrated, in line with De Santis and Spezzaferri (2001), that the same holds for the BF. In the next section the MBF will be introduced, which can be seen as an extension of the BF to multiple populations which avoids this form of inconsistency.

Example 1: Comparison of Two Independent Means. Consider the following simple model:

$$y_i = \theta_1 D_{1i} + \theta_2 D_{2i} + \epsilon_i \text{ with } \epsilon_i \sim \mathcal{N}(0, \omega), \tag{9}$$

where D_{1i} equals 1 for $i = 1, ..., N_1$ and 0 otherwise, D_{2i} equals 1 for $i = N_1 + 1, ..., N_1 + N_2$ and 0 otherwise (and, consequently, θ_1 and θ_2 denote the means in Group 1 and Group 2, respectively, and ω the residual variance), and N_1 and N_2 denote the sample sizes of Groups 1 and 2, respectively, with $N = N_1 + N_2$. Connecting this notation to that of the previous section renders: $\mathbf{Y} = \mathbf{y}$ and $\mathbf{X} = [\mathbf{D}_1, \mathbf{D}_2]$.

Consider testing of $H_1: \theta_1 = \theta_2$ versus $H_c: \theta_1 \neq \theta_2$. Note that the marginal likelihood of H_c is equal to the marginal likelihood of the unconstrained hypothesis $H_u: \theta_1, \theta_2$ because $\theta_1 = \theta_2$ has zero probability assuming a bivariate normal prior for θ_1, θ_2 under H_u . For the exposition that follows we aribitrarily assume that $\hat{\omega} = 1$. The approximated unconstrained posterior and prior distribution of θ_1 and θ_2 from Equations 4 and 6 are then given by

$$g_u(\theta_1, \theta_2 \mid \boldsymbol{y}, \boldsymbol{D}_1, \boldsymbol{D}_2) \approx \mathcal{N}\left(\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix}, \begin{bmatrix} 1/N_1 & 0 \\ 0 & 1/N_2 \end{bmatrix} \right),$$
 (10)

and

$$h_u(\theta_1, \theta_2 \mid [\boldsymbol{y}, \boldsymbol{D}_1, \boldsymbol{D}_2]^b) = \mathcal{N}\left(\begin{bmatrix} 0\\0 \end{bmatrix}, \begin{bmatrix} 1/b \times 1/N_1 & 0\\0 & 1/b \times 1/N_2 \end{bmatrix} \right), \quad (11)$$

respectively, where $b = J^*/N = 1/N$. Note that, with respect to H_1 , the prior means for θ are in agreement with Equation 5.

If we write $\delta = \theta_1 - \theta_2$, then the BF is given by the Savage-Dickey density ratio (Dickey, 1971; Mulder et al., 2010; Wetzels et al., 2010)

$$BF_{1u} = \frac{f_1}{c_1} = \frac{g_u(\delta = 0 \mid \boldsymbol{y}, \boldsymbol{D}_1, \boldsymbol{D}_2)}{h_u(\delta = 0 \mid [\boldsymbol{y}, \boldsymbol{D}_1, \boldsymbol{D}_2]^b)} = \frac{\mathcal{N}(0 \mid \hat{\delta}, N_1^{-1} + N_2^{-1})}{\mathcal{N}(0 \mid 0, (N_1 + N_2)(N_1^{-1} + N_2^{-1}))}.$$

Lets us first of all consider the situation in which N_1 and N_2 go to ∞ with the same rate, that is, let $N_g = a_g n$, for some positive constant a_g , g = 1 or 2 and let $n \to \infty$. If $\hat{\delta} = 0$, then

$$f_1 = \mathcal{N}(0 \mid 0, N_1^{-1} + N_2^{-1}) \to \infty, \tag{12}$$

if $n \to \infty$ and

$$c_1 = \mathcal{N}(0 \mid 0, (N_1 + N_2)(N_1^{-1} + N_2^{-1}) = \mathcal{N}(0 \mid 0, (a_1 + a_2)(a_1^{-1} + a_2^{-1}),$$
(13)

which is a constant independent of n. Equations 12 and 13 imply that $BF_{1u} \to \infty$ if $n \to \infty$ which is consistent. If $\hat{\delta} \neq 0$,

$$f_1 = \mathcal{N}(0 \mid \hat{\delta}, N_1^{-1} + N_2^{-1}) \to 0,$$
 (14)

if $n \to \infty$ and c_1 remains as in Equation 13. This implies that $BF_{1u} \to 0$ if $n \to \infty$ which is consistent.

Now if we fix N_1 and let $N_2 \to \infty$, then in the limit Equation 14 reduces to:

$$f_1 = \mathcal{N}(0 \mid \hat{\delta}, N_1^{-1}),$$
 (15)

and the middle part of Equation 13 reduces to

$$c_1 = \mathcal{N}(0 \mid 0, \infty) \to 0, \tag{16}$$

if $n \to \infty$. This implies that in the limit $BF_{1u} \to \infty$ also if H_u is true, which is inconsistent behavior.

To get more insight about the (in)consistency, the BF was computed for various numerical examples in Tables 1, 2, and 3. In the case of support for H_0 we set $\hat{\theta}_1 = 0$ and $\hat{\theta}_2 = 0$, and in the case of support for H_u we set $\hat{\theta}_1 = -.35$ and $\hat{\theta}_2 = .35$. In both situations we again let $\hat{\sigma}^2 = 1$. As can be seen in Table 1, when $N_1 = N_2$ and both increase with the same rate, $BF_{1u} \to \infty$ if H_1 is true and $BF_{1u} \to 0$ if H_u is true, that is, the Bayes factor

shows consistent behavior. Table 2 shows that BF_{1u} also shows consistent behavior if both sample sizes increase at the same rate if $N_1 \neq N_2$. However, as can be seen in Table 3, if there is support for H_1 , BF_{1u} increases if N_2 increases while N_1 remains fixed, but if there is support in the data for H_u , BF_{1u} at first decreases but then starts to increase, which implies that the evidence accumulates in the wrong direction. As N_2 keeps increasing BF_{1u} goes to infinity as shown above. This is a simple illustration of inconsistent behavior of the BF where multiple populations are considered while the sample size does not increase for all populations. De Santis and Spezzaferri (2001) show that this behavior can also be observed for the fractional Bayes factor. The problem is caused by the fact that the prior variances of θ_1 and θ_2 are dependent on the sample sizes in both groups because b = 1/N(Table 3). As N_2 increases, the fraction that is used to construct the default prior for θ_1 also goes to zero even though the sample size of Group 2 does not increase. This undesirable property can be avoided using population specific fractions in line with Iwaki (1997), Berger and Pericchi (1998), De Santis and Spezzaferri (1999,2001), and Mulder (2014). In the remainder of this paper it will be elaborated how this can be done for the BF to obtain the MBF for multiple populations. \Box

The Approximate Adjusted Fractional Bayes Factor for Multiple Populations

In this section MBF will be introduced. The developments will be illustrated using the comparison of two independent means. Let g = 1, ..., G where G denotes the number of groups and N_g the corresponding sample sizes. Let $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_g, ..., \boldsymbol{\theta}_G, \boldsymbol{\eta}]$, where $\boldsymbol{\theta}_g$ denotes the structural parameters that are unique to Group g and $\boldsymbol{\eta}$ the structural parameters that are shared by all the groups. Then, in line with De Santis and Spezzaferri (2001), the density of the data of the multiple population model can be factored as:

$$p(\boldsymbol{Y}_{1},...,\boldsymbol{Y}_{G} \mid \boldsymbol{\theta}_{1},...,\boldsymbol{\theta}_{G},\boldsymbol{\eta},\boldsymbol{\omega},\boldsymbol{X}_{1},...,\boldsymbol{X}_{G}) = \prod_{g=1}^{G} p_{g}(\boldsymbol{Y}_{g} \mid \boldsymbol{\theta}_{g},\boldsymbol{\eta},\boldsymbol{\omega},\boldsymbol{X}_{g})^{1-b_{g}} p_{g}(\boldsymbol{Y}_{g} \mid \boldsymbol{\theta}_{g},\boldsymbol{\eta},\boldsymbol{\omega},\boldsymbol{X}_{g})^{b_{g}},$$
(17)

where b_g denotes the fraction of the information in the likelihood for Population g that will be used for the specification of the prior distribution.

Example 1: Continued. The following notation will be used to denote which parts of the data belong to Groups 1 and 2. The subscripts 1 and 2 in $\boldsymbol{y}_1, \boldsymbol{y}_2$, denote data sampled from populations 1 and 2, respectively. Analogously, the second subscript in $\boldsymbol{D}_{11}, \boldsymbol{D}_{12}$, and $\boldsymbol{D}_{21}, \boldsymbol{D}_{22}$, denotes data from populations 1 and 2, respectively. Using this notation, the density of the data for the comparison of two independent means can be factored as:

$$p(\boldsymbol{y}_{1}, \boldsymbol{y}_{2} \mid \theta_{1}, \theta_{2}, \sigma^{2}, \boldsymbol{D}_{11}, \boldsymbol{D}_{12}, \boldsymbol{D}_{21}, \boldsymbol{D}_{22}) =$$

$$p(\boldsymbol{y}_{1} \mid \theta_{1}, \sigma^{2}, \boldsymbol{D}_{11}, \boldsymbol{D}_{21})^{1-b_{1}} f(\boldsymbol{y}_{1} \mid \theta_{1}, \sigma^{2}, \boldsymbol{D}_{11}, \boldsymbol{D}_{21})^{b_{1}} \times$$

$$p(\boldsymbol{y}_{2} \mid \theta_{2}, \sigma^{2}, \boldsymbol{D}_{12}, \boldsymbol{D}_{22})^{1-b_{2}} f(\boldsymbol{y}_{2} \mid \theta_{2}, \sigma^{2}, \boldsymbol{D}_{12}, \boldsymbol{D}_{22})^{b_{2}} \propto$$

$$\prod_{i=1}^{N_{1}} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{1})^{2}}{\sigma^{2}/(1 - b_{1})}) \prod_{i=1}^{N_{1}} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{1})^{2}}{\sigma^{2}/b_{1}}) \times$$

$$\prod_{i=N_{1}+1}^{N_{1}+N_{2}} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{2})^{2}}{\sigma^{2}/(1 - b_{2})}) \prod_{i=N_{1}+1}^{N_{1}+N_{2}} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{2})^{2}}{\sigma^{2}/b_{2}}). \quad (18)$$

The covariance matrix of the parameters in Equation 17 can be obtained as a function of the observed or expected Fisher information matrix (the interested reader is referred to Efron and Hinkley, 1978, for an elaboration of the relative (dis)advantages of both types of information). Using the observed Fisher information this leads to

$$-1 \times \begin{bmatrix} \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \theta_1} & 0 & 0 & \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \omega} \\ 0 & \dots & 0 & \dots & \dots \\ 0 & 0 & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \theta_G} & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \mu} & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \omega} \\ \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \eta} & \dots & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} \\ \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \omega} & \dots & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \omega \partial \omega} \end{bmatrix}^{-1} = \\ -1 \times \begin{bmatrix} \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \mu} & 0 & 0 & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} \end{bmatrix}^{-1} \\ 0 & \dots & 0 & \dots & \dots & \dots \\ 0 & 0 & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \theta_G} & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} \\ \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \eta} & \dots & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} \\ \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \eta} & \dots & \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \eta} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} & + \dots + \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} \\ \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} & \dots & \frac{\partial^2 \log p(.)}{\partial \theta_0 \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} & \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} & + \dots + \frac{\partial^2 \log p(.)}{\partial \eta \partial \omega} \\ \end{bmatrix} \right],$$
(19)

where each second order derivative is be evaluated using $[\hat{\theta}, \hat{\eta}, \hat{\omega}]$, that is, the unconstrained maximum likelihood estimates of the model parameters obtained using the full density of the data from Equation 17. If the expected Fisher information is used, the expected value of each entry in the last part of Equation 19 has to be taken. The corresponding normal approximation of the posterior distribution of the structural parameters is

$$g_u(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_G, \boldsymbol{\eta} \mid \boldsymbol{Y}_1, ..., \boldsymbol{Y}_G, \boldsymbol{X}_1, ..., \boldsymbol{X}_G) \approx \mathcal{N}([\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_G, \hat{\boldsymbol{\eta}}], \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}, \boldsymbol{\eta}}),$$
(20)

that is, the multiple population counterpart of Equation 4.

Note that, $\hat{\Sigma}_{\theta,\eta}$ can be constructed using the observed Fisher information matrix for the parameters of each group:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}_{g},\boldsymbol{\eta},\boldsymbol{\omega}} = -1 \times \begin{bmatrix} \frac{\partial^{2}\log p_{g}(.)}{\partial \theta_{g}\partial \theta_{g}} & \frac{\partial^{2}\log p_{g}(.)}{\partial \theta_{g}\partial \eta} & \frac{\partial^{2}\log p_{g}(.)}{\partial \theta_{g}\partial \eta} \\ \frac{\partial^{2}\log p_{g}(.)}{\partial \theta_{g}\partial \eta} & \frac{\partial^{2}\log p_{g}(.)}{\partial \eta \partial \eta} & \frac{\partial^{2}\log p_{g}(.)}{\partial \eta \partial \omega} \\ \frac{\partial^{2}\log p_{g}(.)}{\partial \theta_{g}\partial \omega} & \frac{\partial^{2}\log p_{g}(.)}{\partial \eta \partial \omega} & \frac{\partial^{2}\log p_{g}(.)}{\partial \omega \partial \omega} \end{bmatrix}^{-1}, \text{ for } g = 1, ..., G, \quad (21)$$

where each second order derivative is be evaluated using $\hat{\theta}, \hat{\eta}, \hat{\omega}$, that is, the maximum likelihood estimates of the model parameters obtained using the full density of the data

displayed in Equation 17, that is, *not* only using the data for Group g. Analogously, Equation 21 can be replaced by the corresponding expected Fisher information matrix. Comparing Equation 21 for g = 1, ..., G to Equation 19 shows that the former contains all the elements needed to construct the latter. This is important since the input for **Bain** consists of the covariance matrices per group from which **Bain** constructs the overall covariance matrix. As will be elaborated in the next paragraph, these group specific covariance matrices are needed in order to be able to construct the prior distribution based on a fraction b_g of the information of the data in each group.

Once $\hat{\Sigma}_{\boldsymbol{\theta}_g,\boldsymbol{\eta},\boldsymbol{\omega}}$ for g = 1, ..., G has been obtained it is straightforward to obtain the multiple population counterpart of the prior distribution displayed in Equation 6 which is based on a covariance matrix using a fraction b_g of the information in $\boldsymbol{Y}_g, \boldsymbol{X}_g$ for g = 1, ..., G (see, Equation 17). Using the mathematical rule that $\partial^2 \log p(v, w)^u \partial v \partial w = u \partial^2 \log p(v, w) \partial v \partial w$, it can be seen that:

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}_{g},\boldsymbol{\eta},\boldsymbol{\omega}}^{bg} = -1 \times \begin{bmatrix} \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \theta_{g}}^{bg} & \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \eta}^{bg} & \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \omega}^{bg} \\ \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \eta}^{bg} & \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \eta}^{bg} & \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \omega}^{bg} \\ \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \omega}^{bg} & \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \omega}^{bg} & \frac{\partial^{2} \log p_{g}(.)}{\partial \omega \partial \omega}^{bg} \end{bmatrix}^{-1} = \\ -1 \times \begin{bmatrix} b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \theta_{g}} & b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \eta} & b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \omega} \\ b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \omega} & b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \omega} & b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \omega} \\ b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \theta_{g} \partial \omega} & b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \omega} & b_{g} \frac{\partial^{2} \log p_{g}(.)}{\partial \eta \partial \omega} \end{bmatrix}^{-1} . \end{aligned}$$
(22)

Reassembling these matrices (cf. Equation 19) renders:

$$\hat{\Sigma}^{b}_{\boldsymbol{ heta},\boldsymbol{\eta},\boldsymbol{\omega}} =$$

$$-1 \times \begin{bmatrix} b_1 \frac{\partial^2 \log p_1(.)}{\partial \theta_1 \partial \theta_1} & 0 & 0 & b_1 \frac{\partial^2 \log p_1(.)}{\partial \theta_1 \partial \eta} & b_1 \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \omega} \\ 0 & \dots & 0 & \dots & \dots \\ 0 & 0 & b_G \frac{\partial^2 \log p_G(.)}{\partial \theta_G \partial \theta_G} & b_G \frac{\partial^2 \log p_G(.)}{\partial \theta_G \partial \eta} & b_G \frac{\partial^2 \log p_G(.)}{\partial \theta_G \partial \omega} \\ b_1 \frac{\partial^2 \log p_1(.)}{\partial \theta_1 \partial \eta} & \dots & b_G \frac{\partial^2 \log p_G(.)}{\partial \theta_G \partial \eta} & b_1 \frac{\partial^2 \log p_1(.)}{\partial \eta \partial \eta} + \dots + b_G \frac{\partial^2 \log p_G(.)}{\partial \eta \partial \eta} & b_1 \frac{\partial^2 \log p_1(.)}{\partial \eta \partial \omega} + \dots + b_G \frac{\partial^2 \log p_G(.)}{\partial \eta \partial \omega} \\ b_1 \frac{\partial^2 \log p(.)}{\partial \theta_1 \partial \omega} & \dots & b_G \frac{\partial^2 \log p(.)}{\partial \theta_G \partial \omega} & b_1 \frac{\partial^2 \log p_1(.)}{\partial \eta \partial \omega} + \dots + b_G \frac{\partial^2 \log p_G(.)}{\partial \eta \partial \omega} & b_1 \frac{\partial^2 \log p_1(.)}{\partial \omega \partial \omega} + \dots + b_G \frac{\partial^2 \log p_G(.)}{\partial \omega \partial \omega} \end{bmatrix} \end{bmatrix}^{-1}$$

$$(23)$$

where, $\hat{\Sigma}_{\theta,\eta,\omega}^{b}$ denotes a covariance matrix based on fractions $\boldsymbol{b} = [b_1, ..., b_G]$ of the information in the data, rendering the multiple population adjusted fractional prior distribution of the structural parameters

$$h_u(\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_G, \boldsymbol{\eta} \mid [\boldsymbol{Y}_1, ..., \boldsymbol{Y}_G, \boldsymbol{X}_1, ..., \boldsymbol{X}_G]^{\boldsymbol{b}}) = \mathcal{N}([\boldsymbol{\theta}_{B,1}, ..., \boldsymbol{\theta}_{B,G}, \boldsymbol{\eta}_B], \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta},\boldsymbol{\eta}}^{\boldsymbol{b}}).$$
(24)

Note that, $[...]^{\mathbf{b}}$ in Equation 24 denotes a prior distribution based on fractions \mathbf{b} of the information in the data. It can be interpreted as a default prior that contains the information of group specific data fractions, b_g , for the parameters of interest. Note furthermore that, the subscript B in $\boldsymbol{\theta}_{B,1}, ..., \boldsymbol{\theta}_{B,G}, \boldsymbol{\eta}_B$ highlights that the prior means of the structural parameters are in agreement with Equation 5, that is, centered on the boundary of the hypotheses specified.

The MBF is the counterpart of BF in Equation 2 based on the multiple population posterior and prior distributions displayed in Equations 20 and 24:

$$MBF_{ku} = \frac{\int_{\theta \in H_k} g_u(\theta \mid \mathbf{Y}_1, ..., \mathbf{Y}_G, \mathbf{X}_1, ..., \mathbf{X}_G) d\theta}{\int_{\theta \in H_k} h_u(\theta \mid [\mathbf{Y}_1, ..., \mathbf{Y}_G, \mathbf{X}_1, ..., \mathbf{X}_G]^{\mathbf{b}}) d\theta} \approx \frac{\int_{\theta \in H_k} \mathcal{N}(\theta \mid \hat{\theta}, \hat{\Sigma}_{\theta}) d\theta}{\int_{\theta \in H_k} \mathcal{N}(\theta \mid \theta_B, \hat{\Sigma}_{\theta}^{\mathbf{b}}) d\theta} = \frac{f_k}{c_k}.$$
(25)

Example 1: Continued. Estimates of $\theta_1, \theta_2, \sigma^2$ are easy to obtain. It is well known that using the expected Fisher information the counterpart of Equation 20 for the example at hand is

$$\hat{\Sigma}_{\theta_g,\sigma^2} = \begin{bmatrix} \hat{\sigma}^2/N_g & 0\\ 0 & -1 \times E(\frac{d^2 \log f(.)}{\partial \sigma^2 \partial \sigma^2})^{-1} \end{bmatrix},$$
(26)

from which, using Equation 19, it is straightforward to obtain that

$$\hat{\Sigma}_{\theta_1,\theta_2} = \begin{bmatrix} \hat{\sigma}^2/N_1 & 0\\ 0 & \hat{\sigma}^2/N_2 \end{bmatrix}.$$
(27)

The counterpart of Equation 24 for the example at hand has $\boldsymbol{\theta}_B = [0, 0]$ and, applying Equation 23

$$\hat{\boldsymbol{\Sigma}}_{\theta_1,\theta_2}^{\boldsymbol{b}} = \begin{bmatrix} \hat{\sigma}^2/b_1 \times 1/N_1 & 0\\ 0 & \hat{\sigma}^2/b_2 \times 1/N_2 \end{bmatrix}.$$
(28)

With respect to the computation of Equation 21 three situations can be distinguished:

Situation 1: The multivariate normal linear model with group specific and joint parameters. In the multivariate normal linear model there are z = 1, ..., Z dependent variables and p = 1, ..., P predictors with regression coefficients β_{pz} (where the predictor attached to a possible intercept is a column of ones):

$$y_{1i} = \beta_{11}x_{1i} + \dots + \beta_{P1}x_{Pi} + \epsilon_{1i}$$

$$\dots$$

$$y_{zi} = \beta_{1z}x_{1i} + \dots + \beta_{Pz}x_{Pi} + \epsilon_{zi},$$

$$\dots$$

$$y_{Zi} = \beta_{1Z}x_{1i} + \dots + \beta_{PZ}x_{Pi} + \epsilon_{Zi}$$

$$(29)$$

where $\epsilon_{1i}, ..., \epsilon_{Zi} \sim \mathcal{N}([0, ..., 0], \mathbf{\Omega})$. Multiple populations arise if two or more of the predictors are used to create groups. Two groups with group specific intercepts are created if, for example, $x_{1i} = 1$ if person *i* is a member of Group 1 and 0 otherwise and $x_{2i} = 1$ if person *i* is a member of Group 2 and 0 otherwise. Group specific regression coefficients can additionally be obtained if, for example, $x_{3i} = x_{*i}x_{1i}$ and $x_{4i} = x_{*i}x_{2i}$ (where x_{*i} denotes a continuous predictor for which group specific regression coefficients are required), that is, the predictor x_{3i} gets a regression coefficient β_{3z} in Group 1 and β_{4z} in Group 2. With Z = 1 the model could be:

$$y_i = \beta_1 + \beta_3 x_{3i} + \epsilon_i, \tag{30}$$

for Group 1, and

$$y_i = \beta_2 + \beta_4 x_{3i} + \epsilon_i, \tag{31}$$

for Group 2 and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

For the multivariate normal linear model,

$$\hat{\Sigma}_{\boldsymbol{\theta}_{g},\boldsymbol{\eta}} = \hat{\boldsymbol{\omega}} \otimes [\boldsymbol{X}_{g}^{t} \boldsymbol{X}_{g}]^{-1}, \qquad (32)$$

where θ_g contains all the group specific coefficients, η contains the joint coefficients, ω is a matrix containing the covariance matrix of the residuals, and all the data for the predictors for Group g is collected in X_g . Using Equation 1, hypotheses with respect to the structural parameters $\theta = vec(B)$, where B is a $Z \times P$ matrix containing the regression coefficients β_{pz} , can be formulated. Later in this paper an analysis of covariance model will be used in Example 2 to illustrate Situation 1. Note that, the R function lm can be used to estimate the parameters of the multivariate normal linear model.

Situation 2: Models with only group specific parameters. When all of the parameters (including $\boldsymbol{\omega}$) in the density of the data are group specific, the covariance matrix in Equation 19 will be block diagonal with one block for each group. Consequently, it is straightforward to use R packages tailored to the statistical model of interest to obtain estimates $\hat{\boldsymbol{\theta}}_g$ for, g = 1, ..., G and, for each group, corresponding covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}_g}$. Note that, this does not apply to the example given under Situation 1 (Equations 30 and 31) because σ^2 was not group specific. This would have applied if in addition to the intercept and regression coefficient σ^2 would have been group specific too.

Situation 3: All other situations. In all other situations R packages can be used to obtain the estimates $\hat{\theta}, \hat{\eta}, \hat{\omega}$, but the equations rendering $\hat{\Sigma}_{\hat{\theta}_g, \eta, \omega}$ based on $\hat{\theta}, \hat{\eta}, \hat{\omega}$ and Y_g, X_g for g = 1, ..., G will either have to be programmed in R or obtained through the use of R packages like numDeriv which provides numerical approximations of second order derivatives based on the log density of the data of the statistical model of interest. Later in this paper a logistic regression will be used in Example 3 to illustrate this situation. For users with limited experience in statistical modeling and R, the third situation will be difficult to handle: the likelihood function of the statistical model at hand has to be formulated and numDeriv has to be used to estimate the covariance matrix *per* group (and not the overall covariance matrix) using the overall groups estimates of the model parameters. Currently, one annotated example (a logistic regression model) is provided at the Bain website. Users requiring support in the context of other models can send an e-mail to the first author of this paper with the request to add additional examples to the website.

Choosing b_q

In the case of one population based on Gu, Mulder, and Hoijtink (2017) the R package Bain uses $b = J^*/N$. The remaining question is how to choose b_g for g = 1, ..., G in the case of multiple populations. If the size of the sample obtained from each population is the same, it should not matter whether BF_{ku} or MBF_{ku} is used, that is, it should hold that $BF_{ku} = MBF_{ku}$. Computation of the covariance matrix displayed in Equation 23 in the situation that $N_1 = ... = N_G$ can be done using $b_1 = ... = b_G = b$. Applying this to the one but last diagonal entry of the covariance matrix renders:

$$b_1 \frac{\partial^2 \log p_1(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}} + \dots + b_G \frac{\partial^2 \log p_G(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}} = b \frac{\partial^2 \log p_1(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}} + \dots + b \frac{\partial^2 \log p_G(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}} =$$

$$b(\frac{\partial^2 \log p_1(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}} + ... + \frac{\partial^2 \log p_G(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}}) = b\frac{\partial^2 \log p(.)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}},$$
(33)

with $b = \frac{J^*}{N_1 + \dots + N_G} = \frac{1}{G} \times J^* \times \frac{G}{N_1 + \dots + N_G}$. Therefore a reasonable choice is

$$b_g = \frac{1}{G} \times J^* \times \frac{1}{N_g}.$$
(34)

This choice abides the concept of a minimal fraction from each population to construct an implicit default prior.

Consistency of the Multiple Population Approximate Adjusted Fractional Bayes Factor

De Santis and Spezzaferri (2001) show that their generalized fractional Bayes factor is consistent if $N \to \infty$, that is, if the N_g 's increase at the same rate (cf. De Santis and Spezzaferri, 2001, Theorem 4.1). Below it will be illustrated, via a continuation of Example 1, that the same holds for the MBF, that is, if $N_g \to \infty$, for all g with the same rate, $MBF_{ku} \to \infty$ (or 0) when H_k (or H_u) is true. A more general discussion of the consistency of the MBF is given in the Appendix. The example below also shows that the MBF avoids the inconsistent behavior shown by the BF when fixing the sample size of one population while letting the sample size of the other population go to infinity.

Example 1, Continued. In Example 1 it was illustrated that the BF shows inconsistent behavior if the sample size of one group is fixed while the sample size of the other group goes to infinity. When the MBF is used, the posterior distribution is unchanged and identical to Equation 10. However, the prior distribution changes from Equation 11 to

$$h_{u}(\theta_{1},\theta_{2} \mid [\boldsymbol{y},\boldsymbol{D}_{1},\boldsymbol{D}_{2}]^{\boldsymbol{b}}) \approx \mathcal{N}\left(\begin{bmatrix} 0\\0\\\end{bmatrix}, \begin{bmatrix} 1/b_{1} \times 1/N_{1} & 0\\0 & 1/b_{2} \times 1/N_{2} \end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} 0\\0\\\end{bmatrix}, \begin{bmatrix} 2&0\\0&2 \end{bmatrix}\right), \qquad (35)$$

because $b_1 = \frac{1}{2} \frac{1}{N_1}$ and $b_2 = \frac{1}{2} \frac{1}{N_2}$. As can be seen, the prior distribution in Equation 35 is independent of N_1 and N_2 . This can be interpreted as the amount of prior information being independent of the sample size, which is a desirable property. Also note, that the prior mean does not depend on the information in the data but is chosen to be in agreement with Equation 5.

The MBF of H_1 against H_u is given by

$$MBF_{1u} = \frac{f_1}{c_1} = \frac{g_u(\delta = 0 \mid \boldsymbol{y}, \boldsymbol{D}_1, \boldsymbol{D}_2)}{h_u(\delta = 0 \mid [\boldsymbol{y}, \boldsymbol{D}_1, \boldsymbol{D}_2]^{\boldsymbol{b}})} = \frac{\mathcal{N}(0 \mid \hat{\delta}, N_1^{-1} + N_2^{-1})}{\mathcal{N}(0 \mid 0, 4)} =$$

$$2(N_1^{-1} + N_2^{-1})^{-\frac{1}{2}} \exp\left\{-\frac{\hat{\delta}^2}{2(N_1^{-1} + N_2^{-1})}\right\} \to 2N_1^{\frac{1}{2}} \exp\left\{-\frac{N_1\hat{\delta}^2}{2}\right\},\tag{36}$$

as $N_2 \to \infty$, where $g_u(.)$ and $h_u(.)$ are obtained based on Equations 10 and 35, respectively. As can be seen, if $\hat{\delta} = 0$ and $n \to \infty$, then $f_1 \to \infty$ and c_1 is constant. This implies that $MBF_{1u} \to \infty$. If $\hat{\delta} \neq 0$ and $n \to \infty$, then $f_1 \to 0$ and c_1 is constant. This implies that $MBF_{1u} \to 0$. Stated otherwise, for $n \to \infty$ MBF_{1u} is consistent. Furthermore, if $N_2 \to \infty$ while N_1 is fixed, then if $\hat{\delta} = 0$, in the limit (see the last term of Equation 36) MBF_{1u} = $2N_1^{\frac{1}{2}}$ which is larger than one, that is correctly expresses support for H_1 . Although for $N_2 \to \infty$ MBF_{1u} does not approach ∞ , this is reasonable behavior and the inconsistent behavior of the BF is avoided. If $\hat{\delta} \neq 0$ the limiting behavior of MBF_{1u} = 8.8, that is, H_1 is supported. This too is reasonable, because both the sample size of Group 1 and the effect size are small and therefore the effect is not convincingly different from zero. If both are larger, for example, $N_1 = 49$ and $\hat{\delta} = .5$, MBF_{1u} = .03, that is, H_u is supported. As is illustrated, the degree support for or against H_1 is based on the sample size and the effect size. This too is reasonable behavior and again the inconsistent behavior of the BF is avoided.

As can be seen in the last two columns in the middle and right hand side panel of Table 1 and Table 2, if both sample sizes are proportionally increasing, both BF_{1u} and MBF_{1u} show consistent behavior in the sense that (M)BF_{1u} $\rightarrow \infty$ if $\theta_1 = \theta_2$ and (M)BF_{1u} $\rightarrow 0$ if $\theta_1 \neq \theta_2$. Note that, as required by our choice of b_g in Equation 34, for equal sample sizes in both groups both Bayes factors are equal (see Table 1).

Furthermore, as can be seen in the last two columns in the middle and right hand side panel of Table 3, if one sample size is fixed and the other is increasing, in contrast to BF_{1u} , MBF_{1u} does not show inconsistent behavior in the sense that MBF_{1u} is monotonically increasing if $\hat{\theta}_1 = \hat{\theta}_2$ and MBF_{1u} is monotonically decreasing if $\hat{\theta}_1 \neq \hat{\theta}_2$. As can be seen, in this situation, when only N_2 is increased, MBF_{1u} converges to the upper bound 6.325 (or .546) when $\theta_1 = \theta_2$ (or $\theta_1 \neq \theta_2$) based on the limit in Equation 36. As can be seen comparing the last number on the last line in Table 1 $(N = 200, N_1 = N_2 = 100)$ with the last number on the one but last line in Table 3 $(N = 210, N_1 = 10, N_2 = 200)$ it makes a huge difference in outcome whether or not the sample sizes are balanced. Evidence in favor of the true hypothesis is larger with balanced than with unbalanced sample sizes. \Box

Example 2: Analysis of Covariance

Consider the following analysis of covariance model:

$$y_i = \theta_1 D_{1i} + \dots + \theta_5 D_{5i} + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \text{ with } \epsilon_i \sim \mathcal{N}(0, \sigma^2),$$
 (37)

where D_{1i} is equal to 1 if person *i* is a member of Group 1 and 0 otherwise, the other dummy variables are defined analogously, and both covariates are centered such that θ_1 through θ_5 denote the covariate adjusted means. Equation 37 can be split into five parts, one for each group:

$$y_i = \theta_g + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \tag{38}$$

for g = 1, ..., G and N_g persons in each of the groups. Note that, $\boldsymbol{\theta}_1 = \theta_1, ..., \boldsymbol{\theta}_5 = \theta_5$, $\boldsymbol{\eta} = [\beta_1, \beta_2], \, \boldsymbol{\omega} = \sigma^2$, and $\boldsymbol{Y}_g = \boldsymbol{y}_g, \, \boldsymbol{X}_g = [\boldsymbol{D}_{gg}, \boldsymbol{x}_{1g}, \boldsymbol{x}_{2g}]$, where the second subscript g denotes that the data correspond to the members of Group g.

Applying Equation 17, the density of the data of this model can be factored as

$$p(m{Y}_1,...,m{Y}_5 \mid m{ heta}_1,...,m{ heta}_5,m{\eta},m{\omega},m{X}_1,...,m{X}_5) =$$

$$p_{1}(\boldsymbol{Y}_{1} \mid \boldsymbol{\theta}_{1}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{X}_{1})^{1-b_{1}} p_{1}(\boldsymbol{Y}_{1} \mid \boldsymbol{\theta}_{1}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{X}_{1})^{b_{1}} \times ... \times p_{5}(\boldsymbol{Y}_{5} \mid \boldsymbol{\theta}_{5}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{X}_{5})^{1-b_{5}} p_{5}(\boldsymbol{Y}_{5} \mid \boldsymbol{\theta}_{5}, \boldsymbol{\eta}, \boldsymbol{\omega}, \boldsymbol{X}_{5})^{b_{5}} \propto \prod_{i=1}^{N_{1}} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{1} - \beta_{1} x_{1i} - \beta_{2} x_{2i})^{2}}{\sigma^{2}/(1 - b_{1})}) \prod_{i=1}^{N_{1}} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{1} - \beta_{1} x_{1i} - \beta_{2} x_{2i})^{2}}{\sigma^{2}/b_{1}}) \times ... \times \prod_{i=N_{4}+1}^{N} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{5} - \beta_{1} x_{1i} - \beta_{2} x_{2i})^{2}}{\sigma^{2}/(1 - b_{5})}) \prod_{i=N_{4}+1}^{N} \exp(-\frac{1}{2} \frac{(y_{i} - \theta_{5} - \beta_{1} x_{1i} - \beta_{2} x_{2i})^{2}}{\sigma^{2}/b_{5}}). \quad (39)$$

Maximum likelihood estimates of the parameters $\theta_1, ..., \theta_5, \beta_1, \beta_2, \sigma^2$ of the analysis of covariance model from Equation 37 can, for example, be obtained using the lm function

from the R package. Subsequently, using a well-known result from the regression literature, the realization of Equation 32 for g = 1, ..., 5 is obtained as

$$\hat{\boldsymbol{\Sigma}}_{\theta_g,\beta_1,\beta_2} = \hat{\sigma}^2 \times [\boldsymbol{X}_g^t \boldsymbol{X}_g]^{-1}, \qquad (40)$$

where $\mathbf{X}_g = [\mathbf{D}_{gg}, \mathbf{x}_{1g}, \mathbf{x}_{2g}]$, where the second subscript indicates to which group the data elements belong. Equation 40 is obtained using the expected Fisher information. Since the expected value of the second order derivatives with respect to, either of $\theta_1, ..., \theta_5, \beta_1, \beta_2$ on the one hand and σ^2 on the other hand, are zero, $\hat{\mathbf{\Sigma}}_{\theta_1,...,\theta_5,\beta_1,\beta_2}$ constructed using Equation 19 based on the expected Fisher information for only these parameters is identical to the corresponding part in $\hat{\mathbf{\Sigma}}_{\theta_1,...,\theta_5,\beta_1,\beta_2,\sigma^2}$ (cf. Equations 26 and 27).

Inverting and multiplication with minus one of $\hat{\Sigma}_{\theta_g,\beta_1,\beta_2}$ for g = 1, 2 renders the Fisher information matrices per group. Using Equation 19 these can be assembled into the overal Fisher information matrix which after inverting and multiplication with minus one renders $\hat{\Sigma}_{\theta_1,\theta_2,\beta_1,\beta_2}$. Modifying Equation 40 according to Equation 22 renders

$$\hat{\boldsymbol{\Sigma}}^{b_g}_{\theta_g,\beta_1,\beta_2} = \hat{\sigma}^2 / b_g \times [\boldsymbol{X}^t_g \boldsymbol{X}_g]^{-1}, \qquad (41)$$

that is, the elements of the expected Fisher information matrix for each Group g. Reassembling these elements using Equation 23 renders $\hat{\Sigma}^{\boldsymbol{b}}_{\theta_1,\theta_2,\beta_1,\beta_2}$, that is, the covariance matrix of the prior distribution.

Example 2 will be finished using data from Stevens (1996, Appendix A) concerning the effect of the first year of the Sesame street series on the knowledge of 240 children in the age range 34 to 69 months. We will use the following variables: \boldsymbol{y} , the knowledge of numbers after watching Sesame street; \boldsymbol{x}_1 , the knowledge of numbers before watching Sesame street; \boldsymbol{x}_2 , a test measuring the mental age of children; and $\boldsymbol{D}_1, ..., \boldsymbol{D}_5$ dummy variables representing the children's background (1=disadvantaged inner city, 2=advantaged suburban, 3=advantaged rural, 4= disadvantaged rural, 5=disadvantaged Spanish speaking). The informative hypotheses of interest are:

$$H_1: \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5; \tag{42}$$

and

$$H_2: \{\theta_2, \theta_3\} > \{\theta_1, \theta_4, \theta_5\}.$$
(43)

Hypothesis one states that the knowledge of numbers after watching Sesame street does not depend on background correcting for initial knowledge and mental age. Hypothesis two states that the advantaged children have a higher knowledge after watching Sesame street than the disadvantaged children.

Table 4 presents the input the R package Bain needs in order to evaluate H_1 and H_2 , that is, estimates of the adjusted means, regression coefficients, and residual variance, and, per group, the covariance matrix for the group specific adjusted mean and both regression coefficients, computed using $\hat{\sigma}^2$ (cf. Equation 40), and the sample size. Table 5 first of all presents the posterior covariance matrix of the structural parameters computed from the group specific covariance matrices using Equation 19. Subsequently, the vector **b** computed using $b_g = 1/5 \times 4/N_g$ for g = 1, ..., G is presented. Note that J^* equals 4 because the number of independent constraints in

$$S_{1} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$
(44)

which specifies the equality constraints in H_1 and

$$R_{2} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix}$$
(45)

which specifies the inquality constraints in H_2 , is equal to 4, that is, the number of independent rows in the combination of S_1 and R_2 is equal to 4. Next, the prior covariance matrix of the structural parameters computed using the group specific covariance matrices and **b** and Equations 22 and 23 is displayed. Finally, MBF_{1u} , MBF_{2u} , and MBF_{12} are presented. As can be seen, the support in the data is 2.21 times larger for H_1 than for H_2 , that is, it is slightly more likely that the gain in knowledge of numbers is equal for advantaged and disadvantaged children than that the gain is larger for the advantaged children. More data would be needed to obtain a more decisive conclusion.

Example 3: Logistic Regression

The example presented in the previous section illustrated how $\hat{\Sigma}_{\theta_g,\eta}$ for g = 1, ..., Gcan be computed if the statistical model at hand is a member of the (multivariate) normal linear model (previously labeled *Situation 1*). In this section it will be illustrated how $\hat{\Sigma}_{\theta_g,\eta}$ for g = 1, ..., G can be obtained for models outside the (multivariate) normal linear modeling framework (previously labeled *Situation 3*) based on the observed Fisher information using the R package numDeriv⁴.

Again using the data from Stevens (1996, Appendix A) a logistic regression model is specified in which \boldsymbol{y} , is a child encouraged to watch Sesame street (0=no, 1=yes), is predicted from gender (D_{1i} equals 1 for a girl and zero otherwise, D_{2i} equals 1 for a boy and zero otherwise,), and centered age \boldsymbol{x} :

$$p(y_i \mid D_{1i}, D_{2i}, x_i, \theta_1, \theta_2, \beta) = \frac{\exp(\theta_1 D_{1i} + \theta_2 D_{2i} + \beta x_i)}{1 + \exp(\theta_1 D_{1i} + \theta_2 D_{2i} + \beta x_i)}, \text{ for } i = 1, ..., N.$$
(46)

The hypothesis of interest is:

$$H_1: \theta_1 > \theta_2, \beta > 0, \tag{47}$$

that is, girls are more encouraged than boys and older children are more encouraged than younger children.

⁴https://cran.r-project.org/web/packages/numDeriv/

The top part of Table 6 presents the input the R package Bain needs in order to evaluate H_1 . Note that, $\hat{\Sigma}_{\theta_1,\beta}$ and $\hat{\Sigma}_{\theta_2,\beta}$ are computed using the observed Fisher information matrix rendered by the R package numDeriv using the data for Group 1 (D_{1g}, x_g) and Group 2 (D_{2g}, x_g) , respectively. In the bottom part of Table 6 the output resulting from Bain is presented. It can be observed that H_1 is not supported by the data with a MBF_{1u} = .53. Note that, for the example at hand, $\hat{\Sigma}_{\theta_1,\theta_2,\beta}$ computed using the observed Fisher information matrix is virtually identical to $\hat{\Sigma}_{\theta_1,\theta_2,\beta}$ computed using the expected Fisher information matrix using the R package glm. However note, that this does not always have to be the case. Researchers preferring the expected Fisher information matrix (but, see Efron and Hinkley, 1978) will have to replace the computations with numDeriv by formulas for the expected Fisher information for logistic regression models (see, for example, McCullagh and Nelder, 1989, pp. 115-117).

Discussion

In this paper the approximate adjusted fractional Bayes factor BF, which is suited for the evaluation of informative hypotheses if data are sampled from one population, has been generalized to the multiple population approximate adjusted fractional Bayes factor MBF, which is suited for the evaluation of informative hypotheses if data are sampled from one or multiple populations. Both BF and MBF are implemented in the R package Bain⁵.

The result is a versatile and generally applicable approach for the evaluation of informative hypotheses by means of the Bayes factor in a wide range of statistical models. However, as elaborated earlier in the paper, there are number of topics that deserve further research. The first topic is which sample sizes are required to obtain an accurate normal approximation of the posterior distribution for a wide range of statistical models. The second topic concerns the choice of \boldsymbol{b} , that is, what are the properties of our proposal and what are potential alternatives (the interested reader is referred to Gu, Hoijtink, and

⁵https://informative-hypotheses.sites.uu.nl/software/bain/

Mulder (2016) for one study into this topic). The third topic is further development of **Bain** such that it is easier for users to deal with, what was previously called, Situation 3, that is, models for which numDeriv or other approaches have to be used to obtain the covariance matrix of the parameters of interest for each of the groups in the data set. The fourth topic is more philosophical in nature. It concerns the question whether there is an intrinsic Bayes factor corresponding to our MBF. The fifth topic concerns a modification of the approach presented in this paper such that it can be applied in variable selection problems (see, for example, O'Hare and Sillanpaa, 2009). The spike-and-slap prior is known to perform well in variable selection problems with sparse data, for example, regression models with a relatively large number of persons to number of predictors ratio, and in which only a few predictors are expected to have a substantial regression coefficient. Spike-and-slab prior based variable selection is currently an exploratory approach. In the future we will consider a more confirmatory approach based on an efficient evaluation of sets of informative hypotheses in which it not only is considered if the regression coefficient is substantial, but also its direction, and (partial) orderings of regression coefficients.

References

- Benjamin, D.J., Berger, J.O., Johannesson, M., Nosek, A., Wagenmakers, E-J., ..., Johnson, V.E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2, 6-10. http://dx.doi.org/10.1038/s41562-017-0189-z
- Berger, J.O. and Pericchi, L.R. (1996). The intrinsic Bayes factor for model selection and prediction. Journal of the American Statistical Association, 91, 109-122. DOI: 10.1080/01621459.1996.10476668
- Berger, J.O., Pericchi, L.R. (1998). On criticisms and comparisons of default Bayes factors for model selection and hypothesis testing. *Proceedings of the Workshop on Model Selection*. Pitagora, Bologna, Italy.

- Berger, J.O. and Pericchi, L.R. (2004). Training samples in objective Bayesian model selection. The Annals of Statistics, 32, 841-869. DOI:10.1214/00905360400000229
- Cohen, J. (1994). The earth is round, p<.05. American Psychologist, 49, 997-1003. DOI:10.1037/0003-066X.49.12.997
- De Santis, F., Spezzaferri, F. (1999). Methods for robust and default Bayesian model comparison: the fractional Bayes factor approach. *Int. Statist. Rev.*, 67 (3), 267–286.
- De Santis, F. and Spezzaferri, F. (2001). Consistent fractional Bayes factor for nested normal linear models. Journal of Statistical Planning and Inference, 97,, 305-321. DOI: 10.1016/S0378-3758(00)00240-8
- Dickey, J. (1971). The weighted likelihood ratio, linear hypotheses on normal location parameters. Annals of Mathematical Statistics, 42, 204-223. http://dx.doi.org/10.1214/aoms/1177693507
- Efron, B. and Hinkley, D.V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, 65, 457-483. https://doi.org/10.1093/biomet/65.3.457
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). Bayesian Data Analysis. Boca Raton: Chapman and Hall/CRC.
- Gu, X., Mulder. J., Dekovic, M., and Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods*, 19, 511-527. DOI: 10.1037/met0000017
- Gu, X. (2016). Bayesian evaluation of informative hypotheses. Doctoral Dissertation. University Utrecht. http://informative-hypotheses.sites.uu.nl/books-and-theses/
- Gu. X., Hoijtink, H., and Mulder, J. (2016). Error probabilities in default Bayesian hypothesis testing. *Journal of Mathematical Psychology*, 72 130-143.

http://dx.doi.org/10.1016/j.jmp.2015.09.001

- Gu, X., Mulder, J., and Hoijtink, H. (2017). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. British Journal of Mathematical and Statistical Psychology. DOI: 10.1111/bmsp.12110
- Hoijtink, H. (2012). Informative Hypotheses. Theory and Practice for Behavioral and Social Scientists. Boca Raton: Chapman and Hall/CRC.
- Ioannides, J.P.A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124. DOI: 10.1371/journal.pmed.0020124
- Iwaki, K. (1997). Posterior expected marginal likelihood for comparison of hypotheses. Journal of Economics Asia University, 21, 105-134.
- Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. Journal of the American Statistical Association, 90, 773-795. DOI: 10.1080/01621459.1995.10476572
- Klugkist, I., Kato, B., and Hoijtink, H. (2005). Bayesian model selection using encompassing priors. *Statistica Neerlandica*, 59, 57-69. DOI: 10.1111/j.1467-9574.2005.00279.x
- Klugkist, I., Laudy, O. and Hoijtink, H. (2005a). Inequality Constrained Analysis of Variance: A Bayesian approach. *Psychological Methods*, 10, 477-493. http://dx.doi.org/10.1037/1082-989X.10.4.477
- KLugkist, I., Laudy, O. and Hoijtink, H. (2005b). Bayesian eggs and Bayesian omelettes: Reply to Stern (2005). Psychological Methods, 10, 500-503. http://dx.doi.org/10.1037/1082-989X.10.4.500
- Klugkist, I., Laudy, O., and Hoijtink, H. (2010). Bayesian Evaluation of Inequality and Equality Constrained Hypotheses for Contingency Tables. *Psychological Methods*, 15, 281-299. http://dx.doi.org/10.1037/a0020137

- Kuiper, R., Klugkist, I., and Hoijtink, H. (2010). A Fortran 90 program for confirmatory analysis of variance. *Journal of Statistical Software*, 34, 1-31. DOI: 10.18637/jss.v034.i08
- Kuiper, R. M. and Hoijtink, H. (2013). A Fortran 90 program for the generalization of the order-restricted information criterion. Journal of Statistical Software, 54, 1-19. http://dx.doi.org/10.18637/jss.v054.i08
- McCullagh, P. and Nelder, J.A. (1989). Generalized Linear Models. Second Edition. London: Chapman and Hall.
- Mulder, J., Klugkist, I., Meeus, W., van de Schoot, A., Selfhout, M., and Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology*, 53, 530-546. http://dx.doi.org/10.1016/j.jmp.2009.09.003
- Mulder, J., Hoijtink, H. and Klugkist, I. (2010). Equality and Inequality Constrained Multivariate Linear Models: Model Selection Using Constrained Posterior Priors. Journal of statistical planning and inference, 140, 887-906. http://dx.doi.org/10.1016/j.jspi.2009.09.022
- Mulder, J., Hoijtink, H., and de Leeuw, C. (2012). BIEMS a Fortran90 program for calculating Bayes factors for inequality and equalily constrained models. *Journal of Statistical Software*, 46, 2. DOI:10.18637/jss.v046.i02
- Mulder, J. (2014). Prior adjusted default Bayes factors for testing (in)equality constrained hypotheses. Computational Statistics and Data Analysis, 71, 448-463. http://dx.doi.org/10.1016/j.csda.2013.07.017
- O'Hagan, A. (1995). Fractional Bayes factors for model comparison (with discussion). Journal of the Royal Statistical Society. Series B, 57, 99-138. http://www.jstor.org/stable/2346088

- O'Hagan, A. (1997) Properties of intrinsic and fractional Bayes factors, *Test*, 6, 101-118. http://dx.doi.org/10.1007/BF02564428
- O'Hara, R.B. and Sillanpaa, M.J. (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4, 85-118. DOI:10.1214/09-BA403
- Raftery, A. (1995). Bayesian model selection in social research. In: P. Marsden (Ed.), Sociological Methodology 1995 (pp. 111-163). Cambridge: Blackwell.
- Robert, C.P. (1997). The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation. New York: Springer.
- Royal, R. (1997). Statistical Evidence. A Likelihood Paradigm. New York: Chapman and Hall/CRC.
- Schwartz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461-464. http://dx.doi.org/10.1214/aos/1176344136
- Silvapulle, M.J. and Sen, P.K. (2004). Constrained Statistical Inference: Order, Inequality and Shape Constraints. London: Wiley.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van der Linde, A. (2001). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B, Statistical Methodology, 64, 583-639. 10.1111/1467-9868.00353
- Stern, H.S. (2005). Model inference or model selection: Discussion of Klugkist, Laudy, and Hoijtink (2005). Psychological Methods, 10, 494-499. http://dx.doi.org/10.1037/1082-989X.10.4.494
- Van de Schoot, R., Hoijtink, H., Romeijn, J.-W., and Brugman, D. (2012). A prior predictive loss function for the evaluation of inequality constrained hypotheses. *Journal of Mathematical Psychology*, 56, 13-23. http://dx.doi.org/10.1016/j.jmp.2011.10.001

- Stevens, J. (1996). Applied Multivariate Statistics for the Social Sciences. Mahwah NJ: Lawrence Erlbaum.
- Wagenmakers E-J. (2007). A practical solution to the pervasive problems of p values. Psychonomic Bulletin and Review, 14, 779-804. DOI:10.3758/BF03194105
- Wainer, H. (1999). One cheer for null hypothesis significance testing. Psychological Methods, 4, 212-213. http://dx.doi.org/10.1037/1082-989X.4.2.212
- Wasserstein, R.L. and Lazar, N.A. (2016). The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 70, 129-133. DOI: 10.1080/00031305.2016.1154108
- Wetzels, R., Grasman, R.P.P.P., and Wagenmakers, E-J. (2010). An encompassing prior generalization of the Savage-Dickey density ratio. *Computational Statistics and Data Analysis*, 54, 2094-2102. http://dx.doi.org/10.1016/j.csda.2010.03.016

Appendix: Further discussion of the consistency of the MBF

We consider two different cases. First we consider the case that hypothesis H_k only contains inequality constraints, and no equality constraints. Second we consider the case where H_k contains (only) equality constraints. We will discuss both $N \to \infty$ and one or more but not all of the $N_g \to \infty$.

As was shown in this paper,

$$MBF_{ku} = \frac{f_k}{c_k} = \frac{\int_{\theta \in H_k} \mathcal{N}(\theta \mid \hat{\theta}, \hat{\Sigma}_{\theta}) d\theta}{\int_{\theta \in H_k} \mathcal{N}(\theta \mid \theta_B, \hat{\Sigma}_{\theta}^b) d\theta}.$$
(48)

If H_k only contains inequality constraints, that is, $H_k : R_k \theta > r_k$, MBF reduces to

$$MBF_{ku} = \frac{\Pr(R_k \boldsymbol{\theta} > r_k | \boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})}{\Pr(R_k \boldsymbol{\theta} > r_k | \boldsymbol{\theta}_B, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{\mathbf{b}}) d\boldsymbol{\theta}},$$
(49)

where the Pr(.)'s denote the probabilities that the posterior and prior distribution, respectively, support H_k . To give one simple example, if the prior distribution of $\theta_1, \theta_2, \theta_3$ has $\boldsymbol{\theta}_B = [0, 0, 0]$ and an identity covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{\boldsymbol{b}}$, then $\Pr(R_k \boldsymbol{\theta} > r_k | \boldsymbol{\theta}_B, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\theta}}^{\boldsymbol{b}}) = 1/6$ because there are six possible orderings of three parameters that each have an equal probability.

Let again $N_g = a_g n$, where the a_g 's represent the relative size of the samples from the G populations and let θ^* denote the true value of θ . If the data support H_k , that is, $\theta^* \in H_k$ then if $n \to \infty$ it holds that $\hat{\theta} \to \theta^*$, and the posterior distribution in the numerator of Equation 48 is increasingly concentrated around $\hat{\theta}$ and consequently $f_k \to 1$. Analogously, if the data do not support H_k , that is, $\theta^* \notin H_k$, $f_k \to 0$. This follows from asymptotic theory, see, for example, Gelman et al. (2004, Chapter 4). The prior distribution in the denominator of Equation 48 is independent of the N_g 's and thus independent of n. As can be seen from the combination of Equations 22 and 34, for each group the second order derivatives (which can for the vast majority of statistical models be written as the sum of N_g contributions) are weighted with $b_g = J^*/G \times 1/N_g$, that is, asymptotically each element of Equation 22 is independent of N_g . Consequently, asymptotically c_k is a constant that is independent of n. This is exemplified by Equation 35.

We now have all the ingredients in place to show that MBF_{kc} where H_c : not H_k is consistent. Note that, due to the complementary nature of H_c , $f_c = 1 - f_k$ and $c_c = 1 - c_k$ and thus that

$$MBF_{kc} = \frac{f_k}{c_k} \times \frac{c_c}{f_c}.$$
(50)

Then if $\boldsymbol{\theta}^* \in H_k$ and $n \to \infty$, $\text{MBF}_{kc} \to 1/c_k \times c_c/0 \to \infty$ and if $\boldsymbol{\theta}^* \notin H_k$ and $n \to \infty$, $\text{MBF}_{kc} \to 0/c_k \times c_c/1 \to 0$, which implies consistency.

Theorem 4.1 from De Santis and Spezzaferri (2001) for the generalized fractional Bayes factor and our elaboration in the context of Example 1 earlier in the paper for the MBF provide evidence for consistency if $H_k : S_k \theta = s_k$. Further evidence is obtained by realizing that each equality constraint, e.g. $\theta = 0$ can be written as an about equality constraint $\theta > -z, \theta < z$ for $z \to 0$. If each equality constraint is rewritten in this manner, the exposition given in the beginning of this section applies to $H_k : S_k \boldsymbol{\theta} = s_k$ and also to $H_k : S_k \boldsymbol{\theta} = s_k, R_k \boldsymbol{\theta} > r_k.$

If $N_g \to \infty$ for some but not all of the *G* groups, an analogous line of reasoning can be used show that MBF shows reasonable behavior. If the data support H_k , that is, $\theta^* \in H_k$ and some of the group sizes increase then the posterior distribution in the numerator of Equation 48 is increasingly concentrated around the parameters corresponding to the groups with increasing group sizes (some of the θ_g^*) and η^* . Consequently f_k will become larger but will not attain its maximum value 1.0. Analogously, if $\theta^* \notin H_k$, f_k will become smaller, but will will not attain its minimum value 0.0. Note that, c_k is a constant irrespective of whether $n \to \infty$ or that some of the group sizes go to infinity. These ingredients can be used to show that the behavior of the MBF is reasonable. Looking at Equation 50 it can be seen that: if $\theta^* \in H_k$ MBF will increase (to a boundary value not to infinity) if some of the group sizes go to infinity; and, if $\theta^* \notin H_k$ MBF will decrease (to a boundary value not to zero). A proof and illustration in the context of a simple model can be found in Example 1.

Investigation of consistent behavior of the One Population Bayes factor (BF) and the Multiple Populations Bayes factor (MBF) in the case of support for H_1 ($\bar{x}_1 = (0,0)$) or for H_u

 $(\bar{x}_2 = (-.35, .35))$ in the case of equal sample sizes for both groups increasing at the same rate.

				BF						MBF		
					$ar{\mathbf{x}}_1$	$ar{\mathbf{x}}_2$					$ar{\mathbf{x}}_1$	$ar{\mathbf{x}}_2$
N_1	N_2	b	$\hat{\mathbf{\Sigma}}_{11}/b$	$\hat{\mathbf{\Sigma}}_{22}/b$	BF_{1u}	BF_{1u}	b_1	b_2	$\hat{\mathbf{\Sigma}}_{11}/b_1$	$\hat{\mathbf{\Sigma}}_{22}/b_2$	MBF_{1u}	MBF_{1u}
10	10	.05	2	2	4.47	1.31	.05	.05	2	2	4.47	1.31
25	25	.02	2	2	7.07	.33	.02	.02	2	2	7.07	.33
50	50	.01	2	2	10.00	.02	.01	.01	2	2	10.00	.02
100	100	.005	2	2	14.14	.00	.005	.005	2	2	14.14	.00

Note that: N_1 and N_2 denotes the sample sizes in Group 1 and 2, Respectively; *b* denotes the fraction of information in the density of the data, and b_1 and b_2 denote the fraction of information in the density of the data for Group 1 and 2, respectively; $\hat{\Sigma}_{11}/b$ and $\hat{\Sigma}_{22}/b$ denote the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 23.

Table 2 $\,$

Investigation of consistent behavior of the One Population Bayes factor (BF) and the Multiple Populations Bayes factor (MBF) in the case of support for H_1 ($\bar{\mathbf{x}}_1 = (0,0)$) or for H_u

 $(\bar{\mathbf{x}}_2 = (-.35, .35))$ in the case of unequal sample sizes for the two groups increasing at the same rate.

			BF					MBF				
					$ar{\mathbf{x}}_1$	$ar{\mathbf{x}}_2$					$ar{\mathbf{x}}_1$	$ar{\mathbf{x}}_2$
N_1	N_2	b	$\hat{\mathbf{\Sigma}}_{11}/b$	$\hat{\mathbf{\Sigma}}_{22}/b$	BF_{1u}	BF_{1u}	b_1	b_2	$\hat{\mathbf{\Sigma}}_{11}/b_1$	$\hat{\mathbf{\Sigma}}_{22}/b_2$	MBF_{1u}	MBF_{1u}
10	50	.017	6	1.2	7.74	1.01	.05	.01	2	2	5.77	.75
25	125	.007	6	1.2	12.25	.07	.02	.004	2	2	9.13	.06
50	250	.003	6	1.2	17.32	.00	.01	.002	2	2	12.91	.00

Note that: N_1 and N_2 denotes the sample sizes in Group 1 and 2, Respectively; *b* denotes the fraction of information in the density of the data, and b_1 and b_2 denote the fraction of information in the density of the data for Group 1 and 2, respectively; $\hat{\Sigma}_{11}/b$ and $\hat{\Sigma}_{22}/b$ denote the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 23.

Investigation of consistent behavior of the One Population Bayes factor (BF) and the Multiple Populations Bayes factor (MBF) in the case of support for H_1 ($\bar{x}_1 = (0,0)$) or for H_u

 $(\bar{x}_2 = (-.35, .35)$ in the case of unequal sample sizes where only group size 2 increases.

				BF						MBF		
					$ar{\mathbf{x}}_1$	$ar{\mathbf{x}}_2$					$ar{\mathbf{x}}_1$	$ar{\mathbf{x}}_2$
N_1	N_2	b	$\hat{\mathbf{\Sigma}}_{11}/b$	$\hat{\mathbf{\Sigma}}_{22}/b$	BF_{1u}	BF_{1u}	b_1	b_2	$\hat{\mathbf{\Sigma}}_{11}/b_1$	$\hat{\mathbf{\Sigma}}_{22}/b_2$	MBF_{1u}	MBF_{1u}
10	10	.05	2	2	4.47	1.31	.05	.05	2	2	4.47	1.31
10	25	.029	3.5	1.4	5.92	1.03	.05	.02	2	2	5.34	.93
10	50	.017	6.0	1.2	7.74	1.01	.05	.01	2	2	5.77	.75
10	100	.009	11	1.13	10.48	1.13	.05	.005	2	2	6.03	.65
10	200	.005	21	1.05	14.94	1.41	.05	.0025	2	2	6.17	.60
10	1000	.001	101	1.01	31.78	2.81	.05	.0005	2	2	6.29	.56

Note that: N_1 and N_2 denotes the sample sizes in Group 1 and 2, Respectively; *b* denotes the fraction of information in the density of the data, and b_1 and b_2 denote the fraction of information in the density of the data for Group 1 and 2, respectively; $\hat{\Sigma}_{11}/b$ and $\hat{\Sigma}_{22}/b$ denote the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 11 and $\hat{\Sigma}_{11}/b_1$ and $\hat{\Sigma}_{22}/b_2$ the prior variance of θ_1 and θ_2 from Equation 23. The numbers in italics are referred to in the text.

A Five Group Analysis of Covariance Model: Input for the R Package Bain

$\hat{ heta_1}$	$\hat{ heta_2}$	$\hat{ heta_3}$	$\hat{ heta_4}$	$\hat{ heta_5}$	$\hat{\beta_1}$	$\hat{\beta}_2$	$\hat{\sigma^2}$
29.16	34.38	28.90	27.12	30.89	0.70	0.05	84.06
N_1	N_2	N_3	N_4	N_5			
60	55	64	43	18			
$\hat{\mathbf{\Sigma}}_{ heta_1,eta_1,eta_2}$				$\hat{\mathbf{\Sigma}}_{ heta_2,eta_1,eta_2}$			
1.62	-0.05	0.05		3.07	-0.01	-0.10	
-0.05	0.02	-0.01		-0.01	0.02	-0.01	
0.05	-0.01	0.01		-0.10	-0.01	0.01	
$\hat{\mathbf{\Sigma}}_{ heta_3,eta_1,eta_2}$				$\hat{\mathbf{\Sigma}}_{ heta_4,eta_1,eta_2}$			
2.32	.07	0.09		2.21	.04	.04	
.07	0.03	-0.01		.04	.03	-0.09	
0.09	-0.01	0.01		.04	-0.01	0.01	
$\hat{\mathbf{\Sigma}}_{ heta_5,eta_1,eta_2}$							
5.47	0.20	-0.20					
0.20	0.09	-0.05					
-0.20	-0.05	0.05					

A Five Group Analysis of Covariance Model: Output from

the R Package Bain

$\hat{\mathbf{\Sigma}}_{ heta_1, heta_2, heta_3, he$	$\theta_4, heta_5, eta_1, eta_2$					
1.45	-0.09	0.03	0.02	-0.04	-0.01	0.01
-0.09	1.96	-0.24	-0.12	0.14	0.01	-0.03
0.03	-0.24	1.46	0.07	-0.07	0.00	0.02
0.02	-0.12	0.07	1.99	-0.03	0.00	0.01
-0.04	0.14	-0.07	-0.03	4.72	0.01	-0.01
-0.01	0.01	0.00	0.00	0.01	0.01	-0.00
0.01	-0.03	0.02	0.01	-0.01	-0.00	0.00
b						
0.013	0.015	0.012	0.019	0.044		
$\hat{\Sigma}^{m{b}}_{_{m{ heta}_1}$, $_{m{ heta}_2}$						
01,02,03,	$\beta_4, \theta_5, \beta_1, \beta_2$					
108.14	-5.52	2.02	0.89	-2.66	-0.85	0.67
-5.52	-5.52 129.41	2.02 -13.40	0.89 -6.77	-2.66 7.90	-0.85 0.46	0.67 -1.82
-5.52 2.02	-5.52 -29.41 -13.40	2.02 -13.40 113.04	0.89 -6.77 4.09	-2.66 7.90 -3.86	-0.85 0.46 0.17	0.67 -1.82 0.85
-5.52 2.02 0.89	-5.52 -29.41 -13.40 -6.77	2.02 -13.40 113.04 4.09	0.89 -6.77 4.09 107.20	-2.66 7.90 -3.86 -1.89	-0.85 0.46 0.17 0.14	0.67 -1.82 0.85 0.41
-5.52 2.02 0.89 -2.66	-5.52 -29.41 -13.40 -6.77 7.90	2.02 -13.40 113.04 4.09 -3.86	0.89 -6.77 4.09 107.20 -1.89	-2.66 7.90 -3.86 -1.89 108.07	-0.85 0.46 0.17 0.14 0.51	0.67 -1.82 0.85 0.41 -0.72
108.14 -5.52 2.02 0.89 -2.66 -0.85	-5.52 129.41 -13.40 -6.77 7.90 0.46	2.02 -13.40 113.04 4.09 -3.86 0.17	0.89 -6.77 4.09 107.20 -1.89 0.14	-2.66 7.90 -3.86 -1.89 108.07 0.51	-0.85 0.46 0.17 0.14 0.51 0.31	0.67 -1.82 0.85 0.41 -0.72 -0.14
108.14 -5.52 2.02 0.89 -2.66 -0.85 0.67	-5.52 -5.52 129.41 -13.40 -6.77 7.90 0.46 -1.82	2.02 -13.40 113.04 4.09 -3.86 0.17 0.85	0.89 -6.77 4.09 107.20 -1.89 0.14 0.41	-2.66 7.90 -3.86 -1.89 108.07 0.51 -0.72	-0.85 0.46 0.17 0.14 0.51 0.31 -0.14	0.67 -1.82 0.85 0.41 -0.72 -0.14 0.17
108.14 -5.52 2.02 0.89 -2.66 -0.85 0.67 MBF _{1u}	$\begin{array}{c} -5.52 \\ 129.41 \\ -13.40 \\ -6.77 \\ 7.90 \\ 0.46 \\ -1.82 \\ \end{array}$	2.02 -13.40 113.04 4.09 -3.86 0.17 0.85 MBF ₁₂	0.89 -6.77 4.09 107.20 -1.89 0.14 0.41	-2.66 7.90 -3.86 -1.89 108.07 0.51 -0.72	-0.85 0.46 0.17 0.14 0.51 0.31 -0.14	0.67 -1.82 0.85 0.41 -0.72 -0.14 0.17

Note that, the number in italics is referred to in the text.

A Two Group Logistic Regression

Input for the R Package Bain						
$\hat{ heta_1}$	$\hat{ heta_2}$	\hat{eta}				
.50	.60	01				
N_1	N_2					
125	115					
$\hat{\mathbf{\Sigma}}_{ heta_1,eta}$		$\hat{\mathbf{\Sigma}}_{ heta_2,eta}$				
.03	.00	.04	00			
.00	.00	00	.00			

Output from the ${\tt R}$ Package ${\tt Bain}$

$\hat{\mathbf{\Sigma}}_{ heta_1, heta_2,eta}$			
0.03	-0.00	0.00	
-0.00	0.04	-0.00	
0.00	-0.00	0.00	

0.008	0.009		
$\hat{\mathbf{\Sigma}}^{m{b}}_{_{ heta_{1}, heta_{2},eta}}$			
4.28	-0.02	0.03	
-0.02	4.39	-0.04	
0.03	-0.04	0.06	
MBF_{1u}			

^{.53}

Note that, the number in italics is referred to in the text. Note furthermore, that $\hat{\Sigma}_{\theta_1,\theta_2,\beta}$ does not change when computed using the expected Fisher information.