Psychological Methods

# Computing Bayes Factors From Data With Missing Values

Herbert Hoijtink, Xin Gu, Joris Mulder, and Yves Rosseel

Online First Publication, July 12, 2018. http://dx.doi.org/10.1037/met0000187

American Psychological Association

Psychological Methods


1082-989X/18/$12.00

http://dx.doi.org/10.1037/met0000187

# Computing Bayes Factors From Data With Missing Values

Herbert Hoijtink
Utrecht University

Xin Gu
University of Liverpool

Joris Mulder
Tilburg University

Yves Rosseel
Ghent University

***Abstract***

The Bayes factor is increasingly used for the evaluation of hypotheses. These may be traditional hypotheses specified using equality constraints among the parameters of the statistical model of interest or informative hypotheses specified using equality and inequality constraints. Thus far, no attention has been given to the computation of Bayes factors from data with missing values. A key property of such a Bayes factor should be that it is only based on the information in the observed values. This article will show that such a Bayes factor can be obtained using multiple imputations of the missing values. After introduction of the general framework elaborations for Bayes factors based on default or subjective prior distributions and Bayes factors based on priors specified using training data will be given. It will be illustrated that the approach proposed can be applied using R packages for multiple imputation in combination with the Bayes factor packages Bain and BayesFactor. It will furthermore be illustrated that Bayes factors computed using a single imputation of the data are very inaccurate approximations of the correct Bayes factor.

***Translational Abstract***

The Bayes factor is increasingly used for the evaluation of hypotheses. It is a quantification of the relative support in the data for a pair of hypotheses. If, for example, BF0a = 5, the support in the data is 5 times larger for the null-hypothesis H0 than for the alternative hypothesis Ha. Besides these traditional hypotheses that are specified using equality constraints among the parameters of the statistical model of interest, informative hypotheses specified using equality and inequality constraints can be evaluated using the Bayes factor. A simple example is Hi: m1 > m2 > m3, which states that three means are ordered from largest to smallest. So far, no attention has been given to the computation of the Bayes factor from data with missing values, that is, variables for which a score is missing because a person did not respond or dropped out of the research project. This article will show that such a Bayes factor can be obtained using multiple imputations of the missing values. First of all, using a model for the missing values, multiple data matrices are created in which the missing values are imputed. Second, each of these data matrices is used to evaluate the hypotheses of interest using the Bayes factor. Third, these multiple analyses are summarized into one overall result. It will be illustrated that the approach proposed can be applied using R packages for multiple imputation in combination with the Bayes factor packages Bain and/or BayesFactor. Examples provided in the article concern a simple one-group model, multiple regression, and factor analysis.

*Keywords:* Bayes Factor, informative hypotheses, missing data, multiple imputation

*Supplemental materials:* http://dx.doi.org/10.1037/met0000187.supp



The Bayes factor (Kass & Raftery, 1995; Mulder & Wagenmakers, 2016) is increasingly used for the evaluation of traditional hypotheses specified using equality constraints among the parameters of the statistical model of interest and informative hypotheses (Hoijtink, 2012) specified using inequality and equality constraints (see for applications, e.g., Van Schie, Van Veen, Engelhard, & Klugkist, 2016, and Kolkman, Hoijtink, Kroesbergen, & Leseman, 2013). This article will clarify how these hypotheses can be eval-

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; Yves Rosseel, Department of Data Analysis, Ghent University.

Herbert Hoijtink 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 024.001.003). Xin Gu is supported by NWO-VIDI grant 452-17-006. There has been no prior dissemination of the ideas expressed in this article before it was submitted for publication to *Psychological Methods*.

Correspondence concerning this article should be addressed to Herbert Hoijtink, Department of Methodology and Statistics, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, the Netherlands. E-mail: H.Hoijtink@uu.nl

uated using Bayes factors computed from data with missing values. First of all, multiple imputation of the missing values will be used to create multiple completed data matrices (Rubin, 1987; Schafer, 1997; Van Buuren, 2012). Second, it will be shown that these completed data matrices can be used to approximate the posterior and prior distributions of the parameters of the model at hand such that only the information in the observed values and not the information in the imputed values, is used. Finally, it will be shown that the Bayes factor using only the information in the observed values is a function of these prior and posterior distributions. A strong point of this approach is that each completed data matrix can easily be analyzed using software (e.g., for parameter estimation or Bayes factor computation) that requires data without missing values, and, as will be elaborated in this article, subsequent combination of these analyses renders Bayes factors using only the information in the observed values and not the information in the imputed values.

The focus will be on two approaches for which software implementations are readily available. The first combines multiple imputation with the evaluation of informative hypotheses using the approximate adjusted fractional Bayes factor (Gu, 2016; Gu, Hoijtink, & Mulder, 2016; Gu, Mulder, & Hoijtink, in press; Gu, Mulder, Dekovic, & Hoijtink, 2014; Mulder, 2014) as implemented in the R package Bain (https://informative-hypotheses.sites.uu.nl/software/bain/). The second combines multiple imputation with the g-prior based Bayes factors implemented in the BayesFactor package (http://bayesfactorpcl.r-forge.r-project.org/). This Bayes factor is mainly suited for the evaluation of hypotheses specified using equality constraints (see, e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009) although there are options to test one-sided hypotheses and about equality constraints (Morey & Rouder, 2011). More information about the models that can be handled with both packages will be provided in the section about informative hypotheses. It will be argued that the combination of multiple imputation with Bain or BayesFactor provides a versatile tool for the evaluation of informative hypotheses if the data contain missing values. There are many R packages that can be used to create multiple imputations (see, http://stefvanbuuren.nl/mi/Software.html, for an overview). This article will be illustrated using the R packages norm (Schafer, 1997) which uses, so called, joint multiple imputation, that is, "it is assumed that the data follow a known joint probability distribution with unknown parameters" (Kropko, Goodrich, Gelman, & Hill, 2014) and mice (http://www.stefvanbuuren.nl/mi/mice.html; Van Buuren, 2012) which, like the mi package (https://cran.r-project.org/web/packages/mi/index.html), uses, so-called full conditional imputation; that is, only the distribution of each variable conditional upon the other variables is specified. The differences between and options provided by both approaches will be highlighted later in this article. To ensure applicability of the approaches presented in this article, the data and annotated codes used for each of the examples are made available as online supplementary materials.

Additionally, in Appendix A, there is attention for Bayes factors that can currently not easily be applied in the context of missing data. The fractional Bayes factor (O'Hagan, 1995) can be used for the evaluation of hypotheses specified using equality constraints. It will be presented but not illustrated because it is not implemented in software. Two Bayes factors using minimal training samples (Berger & Pericchi, 1996, 2004; Perez & Berger, 2002) to specify the prior distribution will be presented. The first can be used to evaluate hypotheses specified using equality constraints. The second, implemented in the software package BIEMS (Mulder, Hoijtink, & de Leeuw, 2012), can be used for the evaluation of hypotheses specified using equality and inequality constraints. For both it will be shown how they can be computed from data with missing values. This will not be illustrated because no software exists and BIEMS does not provide the information necessary to computed the Bayes factor from data with missing values. The article is concluded with a short discussion.

## Analysis Model and Imputation Model

In this section the analysis model and the imputation model will be introduced (Meng, 1995; Schafer, 1999). Here and in the following sections the elaborations will be general and illustrated using a normal linear regression model.

To make inferences from data, an analysis model has to be specified. This can be, for example, a normal linear regression model, a structural equation model, or a multilevel model. The data will be denoted by $\boldsymbol{X}$. Often $\boldsymbol{X}$ is a $N \times P$ matrix, where $N$ denotes the number of persons and $P$ the number of variables. In many cases $\boldsymbol{X}$ contains missing values, that is, $\boldsymbol{X}$ can be split into $\boldsymbol{X}_m$ and $\boldsymbol{X}_o$, where $\boldsymbol{X}_o$ contains the observed values and $\boldsymbol{X}_m$ the missing values. The density of the data of the analysis model will be denoted by $p(\boldsymbol{X} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes the parameters of the analysis model. These can be, for example, means, regression coefficients, or factor loadings. In the next section, the model parameters will be decomposed into $\boldsymbol{\theta} = [\gamma, \omega]$ where $\gamma$ denotes the parameters with respect to which hypotheses are formulated, and $\omega$ the nuisance parameters.

### Example 1: Normal Linear Regression

Let $\boldsymbol{X} = [\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2]$, with $\boldsymbol{y} = [y_1, \ldots, y_i, \ldots, y_N]$, $\boldsymbol{x}_1 = [x_{11}, \ldots, x_{i1}, \ldots, x_{N1}]$, and $\boldsymbol{x}_2 = [x_{12}, \ldots, x_{i2}, \ldots, x_{N2}]$ where each variable is continuous. The normal linear regression model is

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + e_i, \text{and } x_{i1}, x_{i2} \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \qquad (1)$$

where $i = 1, \ldots, N$, $\alpha_0$ denotes the intercept, $\alpha_1$ and $\alpha_2$ relate the continuous predictors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to the dependent variable $\boldsymbol{y}$, $e_i \sim \mathcal{N}(0, \sigma^2)$ denotes the error in prediction, and $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_x$ denote the mean vector and covariance matrix of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, respectively. The density of the data is

$$\begin{aligned} p(\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2 \mid \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}x) &= p(\boldsymbol{y} \mid \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{x}_1, \boldsymbol{x}_2 \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \\ &= \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \alpha_0 - \alpha_1 x_{i1} - \alpha_2 x_{i2})^2\right) \\ &\quad \times \mathcal{N}(\boldsymbol{x}_1, \boldsymbol{x}_2 \mid \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x), \end{aligned} \qquad (2)$$

in which a model is specified for the predictors too, that is, they are not assumed to be fixed. In this example $\boldsymbol{\gamma} = [\alpha_1, \alpha_2]$ because these are the parameters with respect to which hypotheses will be formulated in the next section. Consequently, the nuisance parameters are $\omega = [\alpha_0, \sigma^2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x]$.

**End Example 1**

When data contain missing values, $p(\boldsymbol{X}_m, \boldsymbol{X}_o \mid \boldsymbol{\theta})$ cannot straightforwardly be evaluated. This problem can be solved using multiple

imputation (Rubin, 1987; Schafer, 1997; Van Buuren, 2012). Classical applications of multiple imputation consist of three steps. In the first step an appropriately chosen imputation model is used to create multiple completed data matrices in which the missing values are imputed. In the second step each completed data matrix is used to estimate the parameters of the analysis model. In the third step, the Rubin rules (see, e.g., Van Buuren, 2012, pp. 37–38) are used to combine these estimates and their covariance matrix such that the resulting overall estimates and covariance matrix are based only on the information in the observed values and not on the information in the imputed values.

As will be elaborated later in this article and Appendix A, Steps 2 and 3 are only used when the Bayes factor implemented in Bain is computed from data with missing values. However, for all Bayes factors an imputation model has to be used to create completed data matrices. The imputation model will be denoted by $p(\boldsymbol{X}_m, \boldsymbol{X}_o, \boldsymbol{Z}_m, \boldsymbol{Z}_o \mid \boldsymbol{\theta}, \boldsymbol{\eta})$. Note that $\boldsymbol{Z}$ is a $N \times R$ matrix containing auxiliary variables. Analogous to $\boldsymbol{X}$ it can be split into parts containing the observed and missing values: $\boldsymbol{Z}_o$ and $\boldsymbol{Z}_m$. Note furthermore, that the parameters of the imputation model consist of the parameters $\boldsymbol{\theta}$ of the analysis model augmented with parameters $\boldsymbol{\eta}$ that account for the presence of the auxiliary variables $\boldsymbol{Z}$. Why this is a necessary requirement will be elaborated when simplifying Equations 16 to 18 and Equations 20 to 22.

The imputation model has to be chosen such that conditional on the variables in this model the missing data are believed to be missing at random (MAR, Van Buuren, 2012, pp. 6–8, 31–33), that is, the distribution of missingness does not depend on $\boldsymbol{X}_m, \boldsymbol{Z}_m$ (Schafer & Graham, 2002). The big question is which auxiliary variables $\boldsymbol{Z}$ have to be added to the variables $\boldsymbol{X}$ already in the analysis model to achieve MAR. Unless missingness is planned the mechanism causing the missingness is unknow to the researcher. Schafer and Graham (2002) write "When missingness is beyond the researchers' control, its distribution is unknown and MAR is only an assumption. In general, there is no way to test whether MAR holds in a data set." (p. 152). They also write "Although it is not necessary to have a scientific theory underlying an imputation model, it is crucial for that model to be general enough to preserve effects of interest in later analyses." (p. 167). What is often done in practice is that the variables $\boldsymbol{X}$ from the analysis model are augmented with auxiliary variables $\boldsymbol{Z}$ that are expected to be good predictors of the variables in the analysis model containing missing values. Although there can be a convincing argument with respect to which and how many auxiliary variables to add, there is no way to test whether MAR is achieved, MAR is only an assumption. This holds for classical applications of multiple imputation and also when it is used to compute Bayes factor from data with missing values. For a comprehensive overview of multiple imputation and construction of the imputation model the interested reader is referred to the references given in this section. Their elaborations directly apply to the use of multiple imputation when the goal is to compute Bayes factor from data with missing values.

After the choice of $\boldsymbol{X}$ and $\boldsymbol{Z}$, the imputation model $p(\boldsymbol{X}_m, \boldsymbol{X}_o, \boldsymbol{Z}_m, \boldsymbol{Z}_o \mid \boldsymbol{\theta}, \boldsymbol{\eta})$ has to be specified. There are two dominant approaches. The first is joint modeling, that is, a complete specification of $p(\boldsymbol{X}_m, \boldsymbol{X}_o, \boldsymbol{Z}_m, \boldsymbol{Z}_o \mid \boldsymbol{\theta}, \boldsymbol{\eta})$ as is used in Schafer (1997) and the corresponding R packages norm in which $\boldsymbol{X}$ and $\boldsymbol{Z}$ are modeled using a multivariate normal distribution (Chapters 5 and 6), cat in which $\boldsymbol{X}$ and $\boldsymbol{Z}$ are modeled using a log-linear model (Chapters 7 and 8), and mix in which $\boldsymbol{X}$ and $\boldsymbol{Z}$ are modeled using a general location model (Chapter 9). The second is full conditional specification (Van Buuren, 2012) in which only the conditional density of each variable given the other variables is specified as is implemented in the R packages mice and mi.

The statistical theory underlying joint modeling is crystal clear: sample $\boldsymbol{X}_m$, $\boldsymbol{Z}_m$, $\boldsymbol{\theta}$, and $\boldsymbol{\eta}$ iteratively from a posterior distribution based on $p(\cdot)$ and use the sampled values of $\boldsymbol{X}_m, \boldsymbol{Z}_m$ to create completed data matrices. However, joint modeling may be difficult or impossible in a case of complex data structures (e.g., data requiring multilevel modeling) or when the data not only contain groups and continuous dependent and predictor variables, a situation which is covered by the general location model (Schafer, 1997, Chapter 9), but also variables requiring modeling with, for example, a binomial or Poisson distribution. In contrast, full conditional specification is usually easy and straightforward, but until recently the statistical theory underlying this approach was only poorly understood. However, meanwhile is has become clear that when the conditional specifications are linear, logistic, and multinomial regressions, these are compatible with a restricted general location model (Hughes et al., 2014; Liu, Gelman, Hill, Su, & Kropko, 2014; Seaman & Hughes, 2016). Furthermore, there is ample evidence that full conditional specification renders proper imputations even if there is no compatible joint distribution. The interested reader is referred to Kropko et al. (2014); Liu et al. (2014); van Buuren (2007), Van Buuren et al. (2006); and Zhu and Raghunathan (2015) for further elaborations.

As will be elaborated in the section Imputation of Missing Values, for the developments in this article it is irrelevant whether joint modeling or full conditional modeling is used. It is also irrelevant which of the available R packages is used. The illustrations given will be executed with the R packages norm and mice, that is, one using joint modeling and one using full conditional specification. This will provide a stepping stone for researchers wanting to used other packages of either type.

## Example 1 Continued: Normal Linear Regression

When there is one auxiliary variable, that is, $\boldsymbol{Z} = \boldsymbol{z}$, with $\boldsymbol{z} = [z_1, \ldots, z_i, \ldots, z_N]$, a common imputation model $p(\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{z} \mid \boldsymbol{\theta}, \boldsymbol{\eta})$ in the context of normal linear regression is:

$$\begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \boldsymbol{z} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} m_y \\ m_1 \\ m_2 \\ m_z \end{pmatrix}, \begin{pmatrix} s_y^2 & s_{y1} & s_{y2} & s_{yz} \\ s_{y1} & s_1^2 & s_{12} & s_{1z} \\ s_{y2} & s_{12} & s_2^2 & s_{2z} \\ s_{yz} & s_{1z} & s_{2z} & s_z^2 \end{pmatrix} \right), \tag{3}$$

(Schafer, 1997, p. 157; Van Buuren, 2012, pp. 105–107) where, $\boldsymbol{\eta} = [m_z, s_{1z}, s_{2z}, s_z^2, s_{yz}]$ and $\boldsymbol{\theta} = [m_y, m_1, m_2, s_y^2, s_{y1}, s_{y2}, s_1^2, s_{12}, s_2^2]$. Note that, $\boldsymbol{\alpha}$, $\sigma^2$, $\boldsymbol{\mu}_x$, $\boldsymbol{\Sigma}_x$ are a function of $\boldsymbol{\theta}$, and therefore that the imputation model encompasses the analysis model presented in Equation 1. Note furthermore, as will be highlighted later on, that Equation 3 can be implemented using both joint modeling and full conditional modeling (Hughes et al., 2014).

**End Example 1**

## Informative Hypotheses

An important tool in psychological research is the evaluation of hypotheses with respect to the parameters of the analysis model. In this article we consider the class of linear (in)equality constrained hypotheses also known as informative hypotheses (Hoijtink, 2012, Chapter 1). If $\boldsymbol{\theta} = [\boldsymbol{\gamma}, \boldsymbol{\omega}]$, where $\boldsymbol{\gamma}$ is vector of length $J$ containing the parameters that are used in the hypotheses and $\boldsymbol{\omega}$ contains the parameters that are not used, for $k = 1, \ldots, K$ the hypotheses can be formalized as:

$$H_k : \boldsymbol{S}_k\gamma = \boldsymbol{s}_k, \boldsymbol{R}_k\gamma > \boldsymbol{r}_k, \tag{4}$$

where $\boldsymbol{S}_k$ is a $J \times k_1$ matrix imposing $k_1$ equality constraints on $\gamma$, $\boldsymbol{R}_k$ is a $J \times k_2$ matrix imposing $k_2$ inequality constraints, and $\boldsymbol{s}_k$ and $\boldsymbol{r}_k$ are vectors containing constants of size $k_1$ and $k_2$, respectively. Additionally of interest is $H_u : \boldsymbol{\gamma}$, that is, $H_u : \boldsymbol{\gamma}_1, \ldots, \gamma_J$, an hypothesis without constraints on the parameters $\gamma$. This definition of $H_u$ will be used throughout this article. If $H_1$ through $H_K$ are formulated with respect to the parameters collected in $\gamma$, then $H_u$ is the hypotheses that states that there are no restrictions on $\gamma$. Consequently, $H_1$ through $H_K$ are always nested within $H_u$. As will be shown in the next section, this property is used to represent the Bayes factor in terms of the fit and the complexity of the hypotheses entertained.

The persistent critique on $p$ values (see, e.g., Cumming, 2012 and Wagenmakers, 2007) has led to an increased attention for the Bayes factor as a tool for hypothesis evaluation. The R package BayesFactor allows researchers to evaluate classical null-hypotheses using the Bayes factor in the context of the normal linear model (e.g., ANOVA, multiple regression, $t$ tests) and contingency tables. However, also the null-hypothesis has been criticized. Both Cohen (1994) and Royal (1997) criticize the null-hypothesis for being so specific that it is often hard to imagine a population where it holds. This is nicely summarized in Cohen's (1994) title "The Earth is Round ($p < 05$)" that is, a precise hypothesis can be rejected without using data. Another critique is that the null-hypothesis usually does not represent the theory or expectation a researcher has. As is highlighted in Hoijtink (2012) (see also https://informative-hypotheses.sites.uu.nl/), informative hypotheses are an extension of the classical null-hypothesis (see Equation 4) that can represent the theory or expectation of a researcher. Simple examples are: "The depression level after therapy ($\mu_1$) is smaller than the depression level after medication ($\mu_2$) which in turn is smaller than the depression level in a control group ($\mu_3$)" leading to $H_k : \mu_1 < \mu_2 < \mu_3$; "The larger intelligence and social economic status (with standardized regression coefficients $\beta_1$ and $\beta_2$, respectively), the larger income, but intelligence is the stronger predictor" leading to $H_k : \beta_1 > 0, \beta_2 > 0, \beta_1 > \beta_2$; and, "the effect of the new medication on headache ($\mu_1$) is irrelevantly different from the effect of the old medication ($\mu_2$)" leading to $H_k : |\mu_1 - \mu_2| < d \times s$, where $s$ is an estimate of the within group variance and $d$ denotes Cohen's $d$ for which a value demarcating (ir)relevance can be chosen by the researcher. Informative hypotheses are increasingly used by psychological researchers. An overview can be found at https://informative-hypotheses.sites.uu.nl/publications/applications/. Classical hypotheses and informative hypotheses can be evaluated using the Bayes factor implemented in Bain. It has thus far (see, Gu, 2016) been applied in the context of the multivariate normal linear model, logistic regression, multilevel modeling, confirmatory factor analysis, and structural equation modeling. It can be applied to any model for which a normal approximation of the posterior distribution of the model parameters is reasonable.

### Example 1 Continued: Normal Linear Regression

For the regression model displayed in Equation 1 $\boldsymbol{\gamma} = [\alpha_1, \alpha_2]$, that is, hypotheses are formulated with respect to both regression coefficients but not with respect to the other parameters. Three hypotheses will be used that are specifications of Equation 4:

$$H_1 : \alpha_1 = \alpha_2 = 0, \tag{5}$$

$$H_2 : \alpha_1 > 0, \alpha_2 > 0, \alpha_1 > \alpha_2, \tag{6}$$

and,

$$H_u : \alpha_1, \alpha_2, \tag{7}$$

where $H_1$ specifies that both regression coefficients are zero and $H_2$ specifies that both regression coefficients are larger than zero and that the first is larger than the second. Note that the latter restriction only makes sense if the scale on which both predictors are measured is considered to be comparable. As can be seen, $H_1$ and $H_2$ specify hypotheses with respect to $\alpha_1$ and $\alpha_2$, therefore, $H_u$ is the hypothesis in which there are no restrictions on both regression coefficients: $H_u : \alpha_1, \alpha_2$. Note that, for $H_2$

$$\boldsymbol{R}_2 = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} \tag{8}$$

and $\boldsymbol{r}_2 = [0, 0]$. Only the last two restrictions in Equation 6 are specified because they render the first redundant.

**End Example 1**

## Bayes Factor for the Evaluation of Informative Hypotheses in the Absence of Missing Values

The basic structure of the Bayes factor implemented in BayesFactor, Bain and other types of Bayes factors (see Appendix A) used for the evaluation of hypotheses of the form displayed in Equation 4 is identical. Each requires the specification of the posterior and prior distribution of the parameters of the model at hand. The posterior distribution will be denoted by $g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)$, where the subscript $u$ denotes that it is the posterior distribution under $H_u$. Similarly, the prior distribution will be denoted by $h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)$.

Both the mean and the (co)variance(matrix) of the prior distribution require careful specification and will now shortly be discussed. When testing the hypothesis $H_0 : \gamma = 0$ versus $H_u : \gamma \neq 0$ with data $x_i \sim \mathcal{N}(\gamma, \omega)$, for $i = 1, \ldots, N$, Jeffreys (1961) suggests to use $h_u(\gamma, \omega) = h_u(\gamma)h_u(\omega)$ and to give $h_u(\gamma)$ a mean of zero, that is, centered on the null-value used in $H_0$. Inspired by Jeffreys the ANOVA, regression, and $t$ test functions implemented in BayesFactor also center the prior distribution on the values specified by the null hypothesis, see, for example, Rouder, Speckman, Sun, Morey, and Iverson (2009). Mulder (2014) generalizes this principle to hypotheses of the general form displayed in

Equation 4 using, what he calls, an adjustment of the prior mean that can be obtained as follows. Let

$$\boldsymbol{\gamma}_B \in \{\boldsymbol{S}_k\boldsymbol{\gamma}_B = \boldsymbol{s}_k, \boldsymbol{R}_k\boldsymbol{\gamma}_B = \boldsymbol{r}_k\} \quad \text{for} \quad k = 1, \ldots, K, \tag{9}$$

that is, $\boldsymbol{\gamma}_B$ denotes a value of $\boldsymbol{\gamma}$ on the boundary of all the hypotheses under investigation. Let $\hat{\boldsymbol{\gamma}}$ denote the mean of $h_u(\boldsymbol{\gamma}, \boldsymbol{\omega}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)$ with respect to $\boldsymbol{\gamma}$. Then using $\boldsymbol{\gamma}_{adj} = -\hat{\boldsymbol{\gamma}} + \boldsymbol{\gamma}_B$ ensures that the mean of $h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)$ is equal to $\boldsymbol{\gamma}_B$. Applied to $H_0$, this adjustment will render a prior mean of 0 and applied to the ANOVA, regression, and $t$ test functions implemented in BayesFactor it will also render the values specified by the respective null hypotheses. Often the parametric shape of $h_u(\boldsymbol{\gamma})$ is normal, Cauchy, or $t$. This implies that also the prior (co)variance(matrix) or scale(matrix) has to be specified. As will be elaborated later in this article, for the Bayes factors implemented in Bain and discussed in Appendix A this is done using the a fraction of the information in the data, for the Bayes factors implemented in BayesFactor a subjective specification that is independent of the data is required.

As can be seen in, for example, Hoijtink (2012, pp. 51–52), the Bayes factor is often used to compare $H_k$ with $H_u$. Based on Chib's (1995) representation of the Bayes factor, and the derivation presented in Hoijtink (2012, p. 59, see also Appendix B of this article), the Bayes factor comparing the informative hypothesis $H_k$ with the unconstrained hypothesis $H_u$ can be written as:

$$BF_{ku} = \frac{f_k}{c_k}, \tag{10}$$

were $f_k$ denotes the fit of $H_k$, that is,

$$f_k = \int_{\boldsymbol{\theta} \in H_k} \int_{\boldsymbol{\eta}} g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{\theta} d\boldsymbol{\eta}. \tag{11}$$

If hypotheses are specified using only inequality constraints, this is the proportion of the posterior distribution $g_u(\cdot)$ in agreement with $H_k$. The complexity of $H_k$ is denoted by $c_k$, that is,

$$c_k = \int_{\gamma \in H_k} \int_{\omega} \int_{\eta} h_u(\gamma + \gamma_{adj}, \omega, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) d\gamma d\omega d\boldsymbol{\eta}. \tag{12}$$

If hypotheses are specified using only inequality constraints, this is the proportion of the prior distribution $h_u(\cdot)$ in agreement with $H_k$. In the absence of missing data, after specification of $g_u(\cdot)$ and $h_u(\cdot)$, the Bayes factor displayed in Equation 10 can straightforwardly be computed using the packages BayesFactor and Bain. The Bayes factor $BF_{ku}$ quantifies the relative support in the data for $H_k$ and $H_u$. If, for example, $BF_{ku} = 5$ the support for $H_k$ is five times stronger than the support for $H_u$. From the Bayes factors versus the unconstrained model the Bayes factor of $H_k$ versus $H_{k'}$ can be obtained using $BF_{kk'} = BF_{ku}/BF_{k'u}$.

In general it is not easy to compute Bayes factors. However, the interested reader is referred to Chib (1995) and Chib and Jeliazkov (2001) for a generally applicable procedure that gives stable results. However, these procedures do not have to be applied when Bayes factors are computed using BayesFactor or Bain. The Bayes factors implemented in BayesFactor can exactly be computed via the evaluation of relatively simple formulas (see, Rouder et al., 2009, for an example). The Bayes factors implemented in Bain continues a tradition started by Klugkist, Laudy, and Hoijtink (2005) who estimated $f_k$ and $c_k$ by counting the proportion of samples from the posterior and prior distributions, respectively, in agreement with $H_k$. This idea was improved upon by Mulder et al. (2012) who decomposed the fit into one component for each of the $k_1 + k_2$ constraints used to specify $H_k$ such that very accurate estimates of each component could be obtained even if the number of components is large. This idea was further improved by Gu (2016) and Gu, Mulder, and Hoijtink (in press) who simplified the computation of the fit of the $k_1$ components specified using equality constraints and increased the speed of computation without sacrificing accuracy.

## Bayes Factor for the Evaluation of Informative Hypotheses From Data With Missing Values

To obtain the Bayes factor from Equation 10 from data with missing values, $f_k$ and $c_k$, that is, the fit and complexity computed using only the information in the observed values and not the imputed values, have to be computed. The derivation of the fit from data with missing values starts with the observation that

$$\begin{aligned} g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) &= \int_{\boldsymbol{X}_m \boldsymbol{Z}_m} g_u(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{X}_m, \boldsymbol{Z}_m \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{X}_m d\boldsymbol{Z}_m \\ &= \int_{\boldsymbol{X}_m \boldsymbol{Z}_m} g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o) g(\boldsymbol{X}_m, \boldsymbol{Z}_m \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{X}_m d\boldsymbol{Z}_m \\ &\approx \frac{1}{Q} \sum_{q=1}^{Q} g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_m^q, \boldsymbol{Z}_m^q, \boldsymbol{X}_o, \boldsymbol{Z}_o), \end{aligned} \tag{13}$$

where $\boldsymbol{X}_m^q$, $\boldsymbol{Z}_m^q$ denotes the $q$-th imputation of the missing values obtained by sampling from $g(\boldsymbol{X}_m, \boldsymbol{Z}_m \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) = \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\eta}} g_u(\boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{X}_m, \boldsymbol{Z}_m \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{\eta} d\boldsymbol{\theta}$. Note that the latter can be achieved by iteratively sampling from

$$g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o) \tag{14}$$

and

$$p(\boldsymbol{X}_m, \boldsymbol{Z}_m \mid \boldsymbol{\theta}, \boldsymbol{\eta}, \boldsymbol{X}_o, \boldsymbol{Z}_o), \tag{15}$$

and only retaining the sampled values $\boldsymbol{X}_m^q$ and $\boldsymbol{Z}_m^q$ at each iteration $q$. Using Equation 13, the fit can be computed from data with missing values using

$$f_k = \int_{\boldsymbol{\theta} \in H_k} \int_{\boldsymbol{\eta}} \frac{1}{Q} \sum_{q=1}^{Q} g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m^q, \boldsymbol{Z}_m^q, \boldsymbol{X}_o, \boldsymbol{Z}_o) g_u(\boldsymbol{\eta} \mid \boldsymbol{\theta}, \boldsymbol{X}_m^q, \boldsymbol{Z}_m^q, \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{\theta} d\boldsymbol{\eta}. \tag{16}$$

This result can be simplified to

$$f_k = \int_{\boldsymbol{\theta} \in H_k} \frac{1}{Q} \sum_{q=1}^{Q} g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\theta} = \frac{1}{Q} \sum_{q=1}^{Q} f_k^q, \tag{17}$$

or in terms of $\boldsymbol{\gamma}$

$$f_k = \int_{\boldsymbol{\gamma} \in H_k} \frac{1}{Q} \sum_{q=1}^{Q} g_u(\boldsymbol{\gamma} \mid \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\gamma} = \frac{1}{Q} \sum_{q=1}^{Q} f_k^q, \tag{18}$$

where $f_k^q$ denotes the fit computed from the $q$th imputed data matrix (cf. Equation 11). As can be seen, the fit computed from data with missing values is the average of the fits computed for $Q$ imputed data matrices. Note that, if $H_k$ specifies that one or more

of the $\gamma$'s has a fixed value (e.g., $\gamma_1 = 0$) the integration in Equations 17 and 18 evaluates $g_u(\cdot)$ with these $\gamma$'s fixed at the specified values (e.g., $g_u(\gamma_1 = 0, \gamma_2, \ldots, \gamma_J | \boldsymbol{X}_m^q, \boldsymbol{X}_o)$ integrated with respect to $\gamma_2, \ldots, \gamma_J$). The same holds for the integration in Equations 21 and 22 below.

Similarly, the derivation of the complexity starts with the observation that

$$\begin{aligned} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}, \boldsymbol{\eta} \,|\, \boldsymbol{X}_o, \boldsymbol{Z}_o) &= \int_{\boldsymbol{X}_m, \boldsymbol{Z}_m} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}, \boldsymbol{\eta}, \boldsymbol{X}_m, \boldsymbol{Z}_m \,|\, \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{X}_m d\boldsymbol{Z}_m \\ &= \int_{\boldsymbol{X}_m, \boldsymbol{Z}_m} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}, \boldsymbol{\eta} \,|\, \boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o) \\ &\quad g(\boldsymbol{X}_m, \boldsymbol{Z}_m \,|\, \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{X}_m d\boldsymbol{Z}_m \\ &= \frac{1}{Q} \sum_{q=1}^{Q} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}, \boldsymbol{\eta} \,|\, \boldsymbol{X}_m^q, \boldsymbol{Z}_m^q, \boldsymbol{X}_o, \boldsymbol{Z}_o) \end{aligned} \quad (19)$$

Subsequently, using analogous steps as for the fit, the complexity can be computed as:

$$c_k = \int_{\boldsymbol{\gamma} \in H_k} \int_{\boldsymbol{\omega}} \int_{\boldsymbol{\eta}} \frac{1}{Q} \sum_{q=1}^{Q} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega, \boldsymbol{\eta} \,|\, \boldsymbol{X}_m^q, \boldsymbol{Z}_m^q, \boldsymbol{X}_o, \boldsymbol{Z}_o) d\boldsymbol{\gamma} d\boldsymbol{\omega} d\boldsymbol{\eta}, \quad (20)$$

which can be simplified to

$$c_k = \int_{\boldsymbol{\gamma} \in H_k} \int_{\boldsymbol{\omega}} \frac{1}{Q} \sum_{q=1}^{Q} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega} \,|\, \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\gamma} d\boldsymbol{\omega} = \frac{1}{Q} \sum_{q=1}^{Q} c_k^q, \quad (21)$$

or in terms of $\boldsymbol{\gamma}$

$$c_k = \int_{\boldsymbol{\gamma} \in H_k} \frac{1}{Q} \sum_{q=1}^{Q} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj} \,|\, \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\gamma} = \frac{1}{Q} \sum_{q=1}^{Q} c_k^q, \quad (22)$$

where $c_k^q$ denotes the fit computed from the $q$th imputed data matrix (cf. Equation 12). As can be seen, the complexity computed from data with missing values is the average of the complexities computed for $Q$ imputed data matrices.

Equations 17/18 and 21/22 can be used to compute the Bayes factor $BF_{ku}$ displayed in Equation 10 from data with missing values. For the Bayes factors implemented in BayesFactor and those presented in Appendix A, the following representation will be used:

$$BF_{ku} = \frac{f_k}{c_k} = \frac{\frac{1}{Q} \sum_{q=1}^{Q} f_k^q}{\frac{1}{Q} \sum_{q=1}^{Q} c_k^q}. \quad (23)$$

As can be seen, the Bayes factor computed from data with missing values can be represented as the average of the fits computed from imputed data matrices divided by the average of the complexities computed from imputed data matrices. For the Bayes factor implemented in Bain the following representation will be used (cf. Equation 13):

$$\begin{aligned} BF_{ku} &= \frac{f_k}{c_k} \\ &= \frac{\int_{\boldsymbol{\gamma} \in H_k} \frac{1}{Q} \sum_{q=1}^{Q} g_u(\boldsymbol{\gamma} \,|\, \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\gamma}}{\int_{\boldsymbol{\gamma} \in H_k} \frac{1}{Q} \sum_{q=1}^{Q} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj} \,|\, \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\gamma}} \\ &= \frac{\int_{\boldsymbol{\gamma} \in H_k} g_u(\boldsymbol{\gamma} \,|\, \boldsymbol{X}_o) d\boldsymbol{\gamma}}{\int_{\boldsymbol{\gamma} \in H_k} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj} \,|\, \boldsymbol{X}_o) d\boldsymbol{\gamma}}. \end{aligned} \quad (24)$$

## Imputation of the Missing Values

As was elaborated earlier in this article, imputations of the missing values can be obtained using joint modeling of the variables in the imputation model as is, for example, implemented in the R packages norm, cat, and mix, or full conditional modeling as is, for example, implemented in the R packages mice and mi. In this article all illustrations will use norm and/or mice.

In case of joint modeling imputations should be obtained by iteratively sampling from Equations 14 and 15. An approximation of these imputations can be achieved using a joint posterior distribution proportional to

$$p(\boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o \,|\, \boldsymbol{\theta}, \boldsymbol{\eta}) h(\boldsymbol{\theta}, \boldsymbol{\eta}), \quad (25)$$

where $h(\boldsymbol{\theta}, \boldsymbol{\eta})$ is a standard uninformative prior distribution as is implemented in the R packages norm, cat, and mix. This posterior is asymptotically equal to Equation 14 which is proportional to

$$p(\boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o \,|\, \boldsymbol{\theta}, \eta) h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega, \eta \,|\, \boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o), \quad (26)$$

because for increasing sample size the effect of the different prior distributions disappears.

In case of full conditional modeling as is implemented in the R packages mice and mi, the following sampling scheme is used. Let $\boldsymbol{x}^p$ denote one column from $[\boldsymbol{X}, \boldsymbol{Z}]$ for $p = 1, \ldots, P + R$. It is required to specify a conditional imputation model $p(\boldsymbol{x}^p \,|\, \boldsymbol{\lambda}^p, \boldsymbol{x}^{-p})$ for $p = 1, \ldots, P + R$, in which $\boldsymbol{\lambda}^p$ denotes the parameters of the model used to impute $\boldsymbol{x}^p$. Note that $\boldsymbol{x}^{-p}$ denotes all the columns in $[\boldsymbol{X}, \boldsymbol{Z}]$ except $\boldsymbol{x}^p$. Subsequently, the following procedure is used to create imputations:

Step 1. Assign initial values, denoted by $q = 0$, to $\boldsymbol{X}_m^q$ and $\boldsymbol{Z}_m^q$.

Step 2. Iterate Steps 3 and 4 for $p = 1, \ldots, P + R$ during $q = 1, \ldots, Q$ iterations.

Step 3. Sample $\boldsymbol{\lambda}^{p,q}$ from $g(\boldsymbol{\lambda}^p \,|\, \boldsymbol{x}^{p,q-1}, \boldsymbol{x}^{-p,q-1}) \propto p(\boldsymbol{x}^{p,q-1} \,|\, \boldsymbol{\lambda}^p, \boldsymbol{x}^{-p,q-1}) h(\boldsymbol{\lambda}^p)$ where $h(\cdot)$ is a standard uninformative prior.

Step 4. Sample $\boldsymbol{x}_m^{p,q}$ from $p(\boldsymbol{x}_m^p \,|\, \lambda^{p,q}, \boldsymbol{x}^{1,q}, \ldots, \boldsymbol{x}^{p-1,q}, \boldsymbol{x}^{p+1,q-1}, \ldots, \boldsymbol{x}^{P+R,q-1})$.

Application of this procedure results in $q = 1, \ldots, Q$ imputed data matrices $[\boldsymbol{X}_m^q, \boldsymbol{X}_o, \boldsymbol{Z}_m^q, \boldsymbol{Z}_o]$.

## Example 1 Continued: Normal Linear Regression

When norm is used to impute the missing values the following procedure is used:

Step 1. Assign initial values ($q = 0$) to $\boldsymbol{y}_m^q$, $\boldsymbol{x}_{1m}^q$, $\boldsymbol{x}_{2m}^q$, $\boldsymbol{z}_m^q$.

Step 2. Iterate Steps 3 and 4 for $q = 1, \ldots, Q$

Step 3. Sample $\boldsymbol{\alpha}^q$, $\sigma^{2q}$, $\boldsymbol{\mu}_x^q$, $\boldsymbol{\Sigma}_x^q$ from a posterior distribution proportional to Equation 25 in which $p(\boldsymbol{X}_m, \boldsymbol{Z}_m, \boldsymbol{X}_o, \boldsymbol{Z}_o \,|\, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ is equal to Equation 3 and $h(\boldsymbol{\alpha}, \sigma^2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$ is a standard uninformative prior.

Step 4. Sample $\boldsymbol{y}_m^q$, $\boldsymbol{x}_{1m}^q$, $\boldsymbol{x}_{2m}^q$, $\boldsymbol{z}_m^q$ from $p(\boldsymbol{y}_m^q, \boldsymbol{x}_{1m}^q, \boldsymbol{x}_{2m}^q, \boldsymbol{z}_m^q \,|\, \boldsymbol{\alpha}, \sigma^2, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x)$

The interested reader is referred to Schafer (1997, Chapters 5 and 6) for further elaborations.

When mice are used to impute the missing values the following procedure is used:

Step 1. Assign initial values ($q = 0$) to $\boldsymbol{y}_m^q$, $\boldsymbol{x}_{1m}^q$, $\boldsymbol{x}_{2m}^q$, $\boldsymbol{z}_m^q$.

Step 2. Iterate Steps 3 through 5 for $q = 1, \ldots, Q$.

Step 3. Sample $\boldsymbol{\lambda}^q$ from $g(\boldsymbol{\lambda}^q \,|\, \boldsymbol{y}^{q-1}, \boldsymbol{x}_1^{q-1}, \boldsymbol{x}_2^{q-1}, \boldsymbol{z}^{q-1}) \propto p(\boldsymbol{y}^{q-1} \,|\, \boldsymbol{x}_1^{q-1}, \boldsymbol{x}_2^{q-1}, \boldsymbol{z}^{q-1}, \boldsymbol{\lambda}^q) h(\boldsymbol{\lambda}^q)$, where $p(\cdot)$ is based on

$$y_i^{q-1} = \lambda_0^q + \lambda_1^q x_{i1}^{q-1} + \lambda_2^q x_{i2}^{q-1} + \lambda_3^q z_i^{q-1} + e_i, \text{ with}$$

$$e_i \sim \mathcal{N}(0, \lambda_4^q), \tag{27}$$

and $h(\boldsymbol{\lambda}^q) \propto 1/\lambda_4^q$.

Step 4. Sample each missing value $y_{im}^q$ from $p(y_{im} \,|\, x_{i1}^{q-1}, x_{i2}^{q-1}, z_i^{q-1}, \lambda^q) \sim \mathcal{N}(\lambda_0^q + \lambda_1^q x_{i1}^{q-1} + \lambda_2^q x_{i2}^{q-1}, \lambda_4^q)$.

Step 5. Repeat for $x_{i1,m}$, $x_{i2,m}$, $z_{im}$ steps analogous to Steps 3 and 4 based on $p(\boldsymbol{x}_1^{q-1} \,|\, \boldsymbol{y}^q, \boldsymbol{x}_2^{q-1}, \boldsymbol{z}^{q-1}, \boldsymbol{\lambda}_q)$, $p(\boldsymbol{x}_2^{q-1} \,|\, \boldsymbol{y}^q, \boldsymbol{x}_1^q, \boldsymbol{z}^{q-1}, \boldsymbol{\lambda}_q)$, and $p(\boldsymbol{z}^{q-1} \,|\, \boldsymbol{y}^q, \boldsymbol{x}_1^q, \boldsymbol{x}_2^q, \boldsymbol{\lambda}_q)$, respectively. The interested reader is referred to Van Buuren (2012, Chapter 4) for further elaborations.

**End Example 1 Continued**

## Computing "BayesFactor(s)" From Data With Missing Values

The Bayes factors implemented in the R package BayesFactor are mainly suited for the evaluation of hypotheses specified using equality constraints. Many of these Bayes factors can be computed using Equations 17 and 21 based on

$$g_u(\boldsymbol{\theta} \,|\, \boldsymbol{X}_m, \boldsymbol{X}_o) \propto p(\boldsymbol{X}_m, \boldsymbol{X}_o \,|\, \boldsymbol{\theta}) h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega). \tag{28}$$

Because $h_u(\cdot)$ is a user specified prior distribution that does not depend on the data and therefore does not vary over imputed data sets, Equation 23 can be rewritten as:

$$BF_{ku} = \frac{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q}{\frac{1}{Q}\sum_{q=1}^{Q} c_k^q} = \frac{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q}{c_k} = \frac{1}{Q}\sum_{q=1}^{Q} \frac{f_k^q}{c_k} = \frac{1}{Q}\sum_{q=1}^{Q} BF_{ku}^q, \tag{29}$$

where the fit computed from the $q$-th imputed data matrix

$$f_k^q = \int_{\boldsymbol{\theta} \in H_k} g_u(\boldsymbol{\theta} \,|\, \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\theta} \tag{30}$$

and the complexity

$$c_k = \int_{\boldsymbol{\theta} \in H_k} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \boldsymbol{\omega}) d\boldsymbol{\theta}. \tag{31}$$

Stated otherwise, $BF_{ku}$ can be computed from data with missing values as the average of the corresponding Bayes factors computed for each imputed data set. Note that, Equation 29 implies that it is not possible to compute $BF_{uk}$ as the average of the corresponding $BF_{uk}^q$s:

$$BF_{uk} = \frac{\frac{1}{Q}\sum_{q=1}^{Q} c_k^q}{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q} = \frac{c_k}{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q}, \tag{32}$$

which cannot be rewritten in terms of $BF_{uk}^q$s. Therefore, to ensure transitivity relationships $BF_{uk} = 1/BF_{ku}$ and $BF_{kk'} = BF_{ku}/BF_{k'u}$.

Not all functions in the BayesFactor package use prior distributions that do not depend on the data. Examples for which this does hold are Bayes factors for $t$ tests and ANOVAs (Rouder, Morey, Speckman, & Province, 2012; Rouder et al., 2009) but not Bayes factors for multiple regression (Rouder et al., 2012) because there the data for the predictors is used to scale the prior distribution. The approach presented in this article can only be used with the BayesFactor package if the prior distribution is specified independently of the data.

Note that, as highlighted earlier, the approximate equality in Equation 29 is caused by the fact that the posterior distribution used by, for example, norm or mice to create multiple imputations is only asymptotically equal to Equation 28.

## Computing the Approximate Adjusted Fractional Bayes Factor From Data With Missing Values

The Bayes factor implemented in Bain is suited for the evaluation of hypotheses of the general form displayed in Equation 4, that is, specified using equality and/or inequality constraints. It uses a fraction of the information in the likelihood to specify the prior distribution. When introducing the *unadjusted* fractional Bayes factor, O'Hagan (1995) used the following factorization:

$$p(\boldsymbol{X}_m, \boldsymbol{X}_o \,|\, \boldsymbol{\theta}) = c \times \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \,|\, \boldsymbol{\theta})^{1-b(\boldsymbol{X}_m, \boldsymbol{X}_o)} \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \,|\, \boldsymbol{\theta})^{b(\boldsymbol{X}_m, \boldsymbol{X}_o)}, \tag{33}$$

where $c$ denotes the normalizing constant, and $\ell(\cdot)$ the likelihood function. The idea is to use a fraction $b(\boldsymbol{X}_m, \boldsymbol{X}_o)$ of the information in the likelihood function to implicitly specify a default prior distribution (Gilks, 1995). Usually the fraction $b(\boldsymbol{X}_m, \boldsymbol{X}_o)$ is chosen such that it corresponds to the size of a minimal training sample (Berger & Pericchi, 1996, 2004), that is, in the case of complete data, the smallest sample of size $S$ out of $N$ *complete* observations rendering a proper prior distribution: $b(\boldsymbol{X}_m, \boldsymbol{X}_o) = S/N$.

In line with Mulder (2014) this idea can be used to specify the adjusted fractional Bayes factor as a function of the prior and posterior distribution of $\boldsymbol{\theta}$ under $H_u$ :

$$g_u(\boldsymbol{\theta} \,|\, \boldsymbol{X}_m, \boldsymbol{X}_o) \propto \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \,|\, \boldsymbol{\theta})^{1-b(\boldsymbol{X}_m, \boldsymbol{X}_o)} h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega \,|\, \boldsymbol{X}_m, \boldsymbol{X}_o), \tag{34}$$

where the adjusted fractional prior $h_u(\cdot \,|\, \cdot)$

$$h_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega \,|\, \boldsymbol{X}_m, \boldsymbol{X}_o) \propto \ell(\boldsymbol{X}_m, \boldsymbol{X}_o | \boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega)^{b(\boldsymbol{X}_m, \boldsymbol{X}_o)} h(\boldsymbol{\gamma}, \omega) = g_u(\boldsymbol{\gamma} + \boldsymbol{\gamma}_{adj}, \omega \,|\, \boldsymbol{X}_m, \boldsymbol{X}_o, b(\boldsymbol{X}_m, \boldsymbol{X}_o)), \tag{35}$$

is based on a standard uninformative prior $h(\cdot)$ and an adjusted mean for $\boldsymbol{\gamma}$. Note that, $g_u(. \; . \; . | . \; . \; . \; , \; b(. \; . \; .))$ denotes a prior distribution based on a fraction $b(. \; . \; .)$ of the information in the posterior distribution.

Bain (Gu et al., 2014; Gu, Mulder, & Hoijtink, in press) is an abbreviation of Bayesian inequality constrained hypotheses evaluation. It renders an approximation of the adjusted fractional Bayes factor based on a normal approximation of the posterior and prior distributions of the parameters $\boldsymbol{\gamma}$, that is, a normal approximation of Equation 34 and 35 with $\omega$ integrated out. The posterior distribution of $\boldsymbol{\gamma}$ for each imputed data matrix $g_u(\boldsymbol{\gamma} \,|\, \boldsymbol{X}_m^q, \boldsymbol{X}_o)$ (see Equation 18) can be approximated by a multivariate normal distribution with maximum likelihood estimated mean $\hat{\boldsymbol{\gamma}}^q$ and corresponding covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^q$.

Standard results from the multiple imputation literature imply that the posterior distribution of $\gamma$ is a summary of the estimates and covariance matrices obtained for the $q = 1, \ldots, Q$ imputed data matrices (see, e.g., Van Buuren, 2012, pp. 37–38):

$$g_u(\gamma \,|\, \boldsymbol{X}_o) \approx \mathcal{N}(\boldsymbol{\gamma} \,|\, \overline{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}), \tag{36}$$

where

$$\overline{\boldsymbol{\gamma}} = \frac{1}{Q} \sum_{q=1}^{Q} \hat{\boldsymbol{\gamma}}^q. \tag{37}$$

The covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ is obtained using

$$\overline{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} = \frac{1}{Q} \sum_{q=1}^{Q} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^q \tag{38}$$

and

$$\boldsymbol{B} = \frac{1}{Q-1} \sum_{q=1}^{Q} (\boldsymbol{\gamma}^q - \overline{\boldsymbol{\gamma}})(\boldsymbol{\gamma}^q - \overline{\boldsymbol{\gamma}})^t \tag{39}$$

to obtain

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \overline{\boldsymbol{\Sigma}}_{\boldsymbol{\gamma}} + \left(1 + \frac{1}{Q}\right)\boldsymbol{B}. \tag{40}$$

Consequently, the fit $f_k$ from Equation 18 can be approximated by

$$f_k = \int_{\boldsymbol{\gamma} \in H_k} \mathcal{N}(\boldsymbol{\gamma} \,|\, \overline{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) d\boldsymbol{\gamma}. \tag{41}$$

Equation 35 implies that the complexity $c_k$ from Equation 22 can be computed using

$$c_k = \int_{\boldsymbol{\gamma} \in H_k} \mathcal{N}(\boldsymbol{\gamma} \,|\, \boldsymbol{\gamma}_B, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} / b(\boldsymbol{X}_o)) d\boldsymbol{\gamma}, \tag{42}$$

Note that, due to the missing values the effective sample size of $\boldsymbol{X}_o$ is no longer $N$ but $N_o$. The fraction of missing information $\lambda$ can be used to compute $N_o = N - \lambda N$. To compute the fraction of missing information the following quantities are needed (see, e.g., Van Buuren, 2012, pp. 41–43):

$$\alpha = \left(1 + \frac{1}{Q}\right) \mathrm{tr}(\boldsymbol{B}\boldsymbol{\Sigma}^{-1}) / w, \tag{43}$$

where $w$ denotes the Size of $\boldsymbol{\gamma}$,

$$\nu_{old} = \frac{Q-1}{\alpha^2}, \tag{44}$$

$$\nu_{com} = N - w, \tag{45}$$

$$\nu_{obs} = \frac{\nu_{com} + 1}{\nu_{com} + 3} \nu_{com}(1 - \alpha), \tag{46}$$

$$\nu = \frac{\nu_{old} \nu_{obs}}{\nu_{old} + \nu_{obs}}, \tag{47}$$

leading to

$$\lambda = \frac{\nu + 1}{\nu + 3} \alpha + \frac{2}{\nu + 3}. \tag{48}$$

For the Bayes factor implemented in Bain, $b(\boldsymbol{X}_o) = T/N_o$. Note that, $T$ denotes the number of independent constraints imposed on $\gamma$ in all $K$ hypotheses under consideration, that is, the number of independent rows in $[\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K, \boldsymbol{R}_1, \ldots, \boldsymbol{R}_K]$. To give one example: The number of independent constraints in $H_1 : \gamma_1 = \gamma_2 = \gamma_3$ and $H_2 : \gamma_1 > \gamma_2 > \gamma_3$ equals 2. The approximate adjusted fractional Bayes factor implemented in the package Bain is

$$BF_{ku} = f_k / c_k, \tag{49}$$

that is, Equation 24 with $f_k$ and $c_k$ defined in Equations 41 and 42, respectively. As elaborated before, $BF_{ku}$ can be used to compute $BF_{uk}$ and $BF_{kk'}$.

## Example 1 Continued: Normal Linear Regression

This and the next three sections contain applications of the methodology developed in this article. In this section Example 1 will be finished. The next section will show using a simple one variable example that the Bayes factor computed using the proposed methodology is indeed (approximately) equal to the Bayes factor that results if it is computed using only the observed values in $\boldsymbol{X}_o$ and $\boldsymbol{Z}_o$ (cf. Equation 13 and the subsequent derivation of $f_k$ and $c_k$). The subsequent section will compute Bayes factor for a complete data matrix and compare it to the Bayes factor obtained using listwise deletion, the wrong imputation model, the correct imputation model and the correct imputation model extended with auxiliary variables that are not part of the missing data mechanism. The final application section will illustrate the versatility of the approach proposed by applying it to the evaluation of informative hypotheses in the context of a confirmatory factor model when the data contain missing values.

Example 1 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; $\boldsymbol{x}_2$, the knowledge of letters before watching Sesame Street; and, $\boldsymbol{z}$, a test measuring the mental age of children.

The Sesame Street data do not contain missing values. The following missing data mechanism was used to create missing values:

$$P(y_i = m) = \frac{1}{1 + \exp(-25 + x_{i1} + x_{i2})}; \tag{50}$$

$$P(x_{i1} = m, x_{i2} = m) = \frac{1}{1 + \exp(-32 + z_i)}; \quad (51)$$

and,

$$P(z_i = m) = .10 \text{ for } i = 1, \ldots, N. \quad (52)$$

As can be seen, the probability that $z_i$ is missing is .10. The probability that both $x_{i1}$ and $x_{i2}$ are missing increases if the score on $z_i$ decreases. This reflects the idea that children with a lower mental age are less prone to attend the pretest. The probability that $y_i$ is missing increases with decreasing $x_{i1}$ and $x_{i2}$. This reflects the idea that children with a below average performance on the pretest have a smaller probability of showing up during the posttest. Note that, this is a missing at random (MAR) missing data mechanism because the probability that a score on a variable is missing may depend on the scores on the other variables that are available, but *not* on the missing value itself.

In Table 1, descriptives of the resulting data matrix are displayed. As can be seen, about 25%, 15%, 15%, and 10% of the observations are missing from $\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{z}$, respectively. As can also be seen, the scales on which $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are measured are comparable, which implies that it makes sense to compare their regression coefficients as is done in $H_2$. Using norm and mice we created 1,000 completed data matrices in which the missing values are imputed. At the end of this section the choice for 1,000 imputations will be elaborated.

Summarizing the information in the 1,000 completed data matrices imputed using norm rendered

$$\overline{\boldsymbol{\gamma}} = [.644, .006], \quad (53)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \begin{pmatrix} .011 & -.008 \\ -.008 & .014 \end{pmatrix}, \quad (54)$$

and the fraction of missing information $\lambda = .19$. Based on these quantities, Bain evaluated Equation 49 for $H_1$ and $H_2$ rendering $BF_{1u} = .00$ and $BF_{2u} = 7.94$, that is, there is no support in the data for $H_1$ and $H_2$ is 7.94 times as likely as $H_u$. As expected, the Bayes factors based on imputation using mice where virtually identical: $BF_{1u} = .00$ and $BF_{2u} = 7.97$. The data matrix used and annotated R code detailing the steps in the analyses presented in this section can be found in the online supplementary materials.

This relatively large number of $Q = 1{,}000$ completed data matrices was chosen to ensure a stable estimate of the fraction of missing information λ. In Table 2, λ was computed using three different seeds to start the sampling of the missing values as executed by mice. As can be seen, for $Q = 10$ the λ may differ as much as .12. When $Q$ is increased the differences become smaller. For $Q = 500$ the largest difference is about .02, which ensures that $Q = 1{,}000$ renders accurate estimates of λ. In each example in this article $Q = 1{,}000$ will be used. Researchers applying the approach presented in this article are well-advised to also use a large value for $Q$ and to verify using different seeds for the multiple imputation package used that the value chosen renders a stable estimate of λ.

Table 1
*Descriptive Statistics for the Sesame Street Data With Missing Values*

| Variable | Mean | Minimum | Maximum | Number missing |
|---|---|---|---|---|
| $\boldsymbol{y}$ | 33.84 | 7 | 54 | 57 |
| $\boldsymbol{x}_1$ | 22.52 | 2 | 52 | 37 |
| $\boldsymbol{x}_2$ | 16.58 | 1 | 55 | 37 |
| $\boldsymbol{z}$ | 46.77 | 27 | 99 | 27 |

Table 2
*He Choice of Q and the Stability of* λ

| $Q$ | Seed = 999 | Seed = 1,234 | Seed = 4,321 |
|---|---|---|---|
| 10 | .402 | .276 | .285 |
| 50 | .272 | .291 | .299 |
| 100 | .309 | .278 | .312 |
| 500 | .300 | .289 | .305 |
| 1000 | .308 | .298 | .293 |

**End Example 1 Continued**

## Example 2: Approximate Equality of Bayes Factors Computed From Observed Values and Completed Data Matrices

Consider the analysis model

$$x_i \sim \mathcal{N}(\gamma, \omega) \text{ for } i = 1, \ldots, N, \quad (55)$$

and the informative hypotheses $H_1 : \gamma = 0$ and $H_2 : \gamma > 0$. Note that, there is only one γ and therefore that $H_u : \gamma$ is a hypothesis without restrictions on this gamma. Data will be generated such that the sample average of the observed values in $x$ is $\bar{x} \in \{-.2, 0, .2, .5\}$, and the sample standard deviation of the observed values is $s = 1$, the number of observed values is $N_o = 30$ and $N_m = 20$ missing values are added, that is, the missing values are MAR (in this case in fact missing completely at random). This setup allows for inference with (that is, based on $\boldsymbol{X}_o, \boldsymbol{X}_m$) and without (that is, based on $\boldsymbol{X}_o$) imputation of the missing values.

This setup enables the illustration of what was shown by means of equations (cf. Equation 13 and the subsequent derivation of $f_k$ and $c_k$) earlier in this article: if the missing values are MAR the Bayes factor computed using only the observed values and the Bayes factor computed from multiple completed data matrices are approximately identical. Note that, results are only approximately equal because the posterior distribution used for imputation is only asymptotically equal to the posterior distributions used to compute Bayes factors. Note furthermore, that the approximate equality was shown in general, but can only be illustrated using this simple setup because in more complicated setups like, for example, Example 1, the Bayes factor can be computed using the multiple imputation approach proposed in this article, but there is no way in which the Bayes factor can directly be computed from the observed values like in the simple example presented in this section. Consider, for example, the regression model from Example 1 with missing values in $\boldsymbol{y}$, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and the auxiliary variable $\boldsymbol{z}$ needed to achieve MAR. Neither BayesFactor nor Bain can be used to compute Bayes factors using these data because they require the data to be complete. However, as elaborated earlier in this article, this can be solved based on multiple imputation of the missing values.

Columns 3, 4, and 5 of Table 3 contain the input needed to use Bain for the evaluation of $H_1$ and $H_2$ (cf. Equations 41 and 42): $\overline{\boldsymbol{\gamma}}$; $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$; and, λ. Each of these are estimated based on 1,000 imputations of the missing values using mice. Note that, the imputation model was identical to the analysis model, and therefore the same

Table 3
*Bayes Factors Computed Using Only the Observed Values and Multiple Imputation With the Packages Bain and BayesFactor*

| | | | | | Bain | | BayesFactor | |
|---|---|---|---|---|---|---|---|---|
| | $\bar{x}$ | $\bar{\gamma}$ | $\boldsymbol{\Sigma}_\gamma$ | $\lambda$ | $BF_{1u}$ | $BF_{2u}$ | $BF_{1u}$ | $BF_{2u}$ |
| Observed | −.2 | −.20 | .03 | | 3.01 | .27 | 2.98 | .30 |
| Imputed | −.2 | −.20 | .03 | .45 | 3.03 | .29 | 2.94 | .30 |
| Observed | 0 | .00 | .03 | | 5.47 | 1 | 5.14 | 1 |
| Imputed | 0 | .00 | .03 | .45 | 5.26 | 1 | 5.09 | 1.02 |
| Observed | .2 | .20 | .03 | | 3.01 | 1.72 | 2.98 | 1.70 |
| Imputed | .2 | .20 | .03 | .45 | 2.87 | 1.73 | 2.78 | 1.72 |
| Observed | .5 | .50 | .03 | | .13 | 1.99 | .23 | 1.99 |
| Imputed | .5 | .50 | .03 | .45 | .13 | 1.99 | .19 | 1.99 |



irrespective of whether a joint or full conditional specification is used. As can be seen, the estimates based on analysis of the observed values and multiple imputation are very similar (to two decimal places), but the fraction of missing information $\lambda$ is slightly overestimated. The resulting values of the Bayes factors based on the observed values and multiple imputation (Equation 49) for $BF_{1u}$ (column 5) and $BF_{2u}$ (column 6) are similar. With a better estimate of $\lambda$ they would have been virtually the same. Nevertheless, this simple example shows that the Bayes factor computed from the observed values can adequately be approximated by the imputed data based on Bayes factor because the differences that can be observed do not change the interpretation of the resulting numbers.

Using the function ttestBF from the BayesFactor R package, the Bayes factors $BF_{1u}$ and $BF_{2u}$ (cf. Equation 29) are computed using only the observed values multiple imputation. In this package the analysis model is reparameterized as $x_i \sim \mathcal{N}(\sqrt{\omega}\gamma, \omega)$, where $\gamma$ denotes the standardized effect size, $h_u(\gamma) \sim Cauchy(0, .707)$ where .707 is the prior scale, and $h(\omega) = 1/\omega$. As can be seen in the last two columns of Table 3, there is a close correspondence between the Bayes factors based on the observed values and multiple imputation.

In Figure 1 the $BF^q_{2u}$ for $q = 1, \ldots, 1{,}000$ as computed using the BayesFactor R package are displayed for $\bar{x} = 0$, that is, one data matrix with missing values that was imputed 1,000 times and for each imputed data matrix the Bayes factor was computed. As can be seen, using single imputation to obtain the Bayes factor is not a good idea because the $BF^q_{2u}$ vary between 0 and 2.0. Using the $BF_{2u}$ based on only one of these imputations would render a Bayes factor that is essentially a random number from the interval 0 to 2.0. which can easily be substantially different from the correct value 1.0.

Note that, the codes used to obtain the results in Table 3 and Figure 1 can be found in the online supplementary materials. Note furthermore that, the analyses presented in this section were also executed using $N_o$ equal to 40 and 20 and $N_m$ equal to 10 and 30, respectively. The results were completely in line with those described above.

## Example 3: The Effect of the Choice of the Auxiliary Variables

This section will illustrate the effect of the choice of the auxiliary variables on the Bayes factor computed from data with missing values. One data matrix without missing values is used consisting of a dependent variable $\boldsymbol{y} = [y_1, \ldots, y_i, \ldots, y_N]$, three potential auxiliary variables $\boldsymbol{z}_1 = [z_{11}, \ldots, z_{i1}, \ldots, z_{N1}]$, $\boldsymbol{z}_2 = [z_{12}, \ldots, z_{i2}, \ldots, z_{N2}]$, and $\boldsymbol{z}_3 = [z_{13}, \ldots, z_{i3}, \ldots, z_{N3}]$, and a grouping variable $\boldsymbol{x} = [x_1, \ldots, x_i, \ldots, x_N]$, that attains the value 0 for the members of Group 1 and the value 1 for the members of Group 2. For both groups, the means and covariance matrix of the dependent and auxiliary variables are displayed in Table 4.

The analysis model of interest in this example is

$$y_i = \alpha_0 + \alpha_1 D_i + e_i \text{ with } e_i \sim \mathcal{N}(0, \sigma^2), \tag{56}$$

that is, a two-group ANOVA in which $\alpha_1$ denotes the difference in means between Groups 2 and 1 because $D_i$ equals 0 if person $i$ is a member of Group 1 and 1 if person $i$ is a member of Group 2. The hypotheses of interest are $H_1 : \alpha_1 = 0$ and $H_2 : \alpha_1 > 0$. Missing data were created such that it holds in Group 1 that the

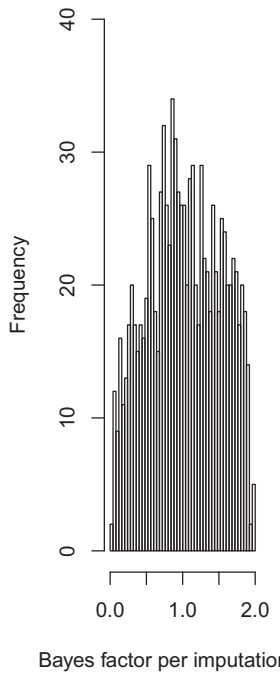


*Figure 1.* The distribution of $BF^q_{2u}$ for $q = 1, \ldots, 1{,}000$ imputations of one data set with missing values for which $\bar{x} = 0$.

Table 4
*Per Group, Means and Covariance Matrix of the Dependent and Auxiliary Variables*

| | Group 1 | | | | | Group 2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variable | Mean | Covariance matrix | | | | Mean | Covariance matrix | | | |
| $\boldsymbol{y}$ | −.20 | .72 | | | | .44 | .91 | | | |
| $\boldsymbol{z}_1$ | −.03 | .18 | .74 | | | −.29 | .32 | .98 | | |
| $\boldsymbol{z}_2$ | −.14 | .40 | .36 | 1.16 | | .02 | .60 | .51 | 1.25 | |
| $\boldsymbol{z}_3$ | −.15 | .32 | .32 | .51 | .80 | −.02 | .57 | .50 | .86 | 1.36 |

smaller $z_{i1}$ the larger the probability that $y_i$ is missing and in Group 2 the larger $z_{i1}$ the larger the probability that $y_i$ is missing:

$$P(y_i = m \mid D_i = 0) = \frac{1}{1 + \exp(z_{i1})}, \tag{57}$$

and

$$P(y_i = m \mid D_i = 1) = \frac{1}{1 + \exp(-z_{i1})}. \tag{58}$$

This rendered 24 missing values in $\boldsymbol{y}$ for both the members of Group 1 and Group 2.

The hypotheses $H_1$ and $H_2$ were evaluated using six scenarios:

1. Using the full data before the creation of missing values.

2. Using the data remaining after list-wise deletion of the cases with missing values.

3. Using the data after multiple imputation of the missing values using $\boldsymbol{y}$ and $\boldsymbol{x}$ without auxiliary variables in a multivariate normal imputation model (cf. Equation 3). Note that, $\boldsymbol{x}$ can be modeled in this manner only because it does not contain missing values. Would it have contained missing values the imputation model should be one of the family of general location models (Schafer, 1997, Chapter 9).

4. Like Scenario 3 but now using $\boldsymbol{y}$, $\boldsymbol{x}$, and the variable predicting the missing values $\boldsymbol{z}_1$.

5. Like Scenario 3 but now using $\boldsymbol{y}$, $\boldsymbol{x}$, $\boldsymbol{z}_1$, and the variables not predicting the missing values $\boldsymbol{z}_2$ and $\boldsymbol{z}_3$.

6. Like Scenario 3 but now using $\boldsymbol{y}, \boldsymbol{x}$, and only the variables not predicting the missing values $\boldsymbol{z}_2$ and $\boldsymbol{z}_3$.

The results are displayed in Table 5. The following can be observed:

1. When the full data are used, the support in the data is substantially larger for $H_2$ than for $H_1$ ($BF_{21} = 102.62$). This is not surprising, because $\hat{\alpha}_1 = .64$ and is standard deviation is $\sqrt{.033} = .18$. Note that, the results obtained in Scenarios 2 through 6 *cannot* straightforwardly be compared with the full data results because here the fraction of missing information equal 0, that is, $N_o = N = 100$, while in the other scenarios the fraction of missing information is substantially larger than 0. This will lead to larger standard deviations for $\hat{\alpha}_1$ and smaller Bayes factors in the other scenarios. What can be learned from the full data is that scenarios that adequately account for the missing data mechanism should render estimates of $\alpha$ close to .64, should render standard deviations larger than .18, and should render Bayes factors favoring $H_2$.

2. Because the missing data mechanism (missingness depends on $\boldsymbol{z}_1$) is not accounted for in case of listwise deletion and imputation using only $\boldsymbol{y}$ and $\boldsymbol{x}$ the missing values are not MAR. This can for Scenarios 2 and 3 clearly be seen from $\hat{\alpha}_1 = .39$ (too small) and a $BF_{21}$ slightly favoring $H_1$ instead of $H_2$. This implies that inadequate representation of the missing data mechanism in the imputation model will render incorrect inferences.

3. As noted earlier, what is often done in practice is construction of an imputation model using the variables from the analysis model augmented with auxiliary variables that are expected to be good predictors of the variables in the analysis model containing missing values. In Scenario 4 the imputation model uses $\boldsymbol{y}$, $\boldsymbol{x}$, and the variable $\boldsymbol{z}_1$ which explains the missing values. As can be seen, with the correct imputation model an accurate estimate $\hat{\alpha}_1 = .61$ is obtained and $BF_{21} = 3.18$ correctly indicates a preference for $H_2$ irrespective of the fact that 52% of the information in the data is missing leading to a standard deviation of $\sqrt{.077} = .28$ which is (and should be) substantially larger than .18 (the standard deviation obtained using the full data).

4. Scenario 5 is an extension of Scenario 4 in which also $\boldsymbol{z}_2$ and $\boldsymbol{z}_3$ are added to the variables in the imputation model. This is in line with Schafer and Graham (2002) who write "Although it is not necessary to have a scientific theory underlying an imputation model [which would imply that researchers *know* that $\boldsymbol{z}_1$ causes the missing values], it is crucial for that model to be general enough to preserve effects of interest in later analyses [that is, use the variables in the analysis model and variables related to the variables containing missing values]." As can be seen, even if "unnecessary" variables are included in the imputation model, an accurate estimate $\hat{\alpha}_1 = .65$ is obtained and $BF_{21} = 6.36$ correctly prefers $H_2$ over $H_1$.

5. Continuing Scenario 5, even if $\boldsymbol{z}_1$ is not part of the imputation model but it does contain variables like $\boldsymbol{z}_2$ and $\boldsymbol{z}_3$ that are related to the variables containing missing

Table 5
*Key Quantities and Bayes Factors for Six Scenarios*

| Scenario | $\hat{\alpha}_1$ | $\boldsymbol{\Sigma}_{\alpha_1}$ | $\lambda$ | $BF_{1u}$ | $BF_{2u}$ | $BF_{21}$ |
|---|---|---|---|---|---|---|
| 1 | .64 | .033 | 0 | .02 | 2 | 102.62 |
| 2 | .39 | .070 | .43 | 2.39 | 1.86 | .78 |
| 3 | .39 | .074 | .51 | 2.48 | 1.85 | .75 |
| 4 | .61 | .077 | .52 | .62 | 1.97 | 3.18 |
| 5 | .65 | .067 | .45 | .31 | 1.98 | 6.36 |
| 6 | .60 | .063 | .43 | .43 | 1.98 | 4.64 |

values, an adequate estimate $\hat{\alpha}_1 = .60$ is obtained, and $BF_{21} = 4.64$ expresses and preference for $H_2$. This is in line with ". . . that [imputation] model to be general enough to preserve effects of interest . . ." (Schafer & Graham, 2002, p. 167), that is, the imputation model does not necessarily have to contain the variables causing the missing values.

In summary, even if the missing data mechanism is unknown, using the variables in the analysis model extended with variables related to the variables containing missing values, may very well render adequate estimates of the effects of interest and Bayes factors adequately expressing preference for the hypotheses of interest. However, although researchers can (and should) argue in favor of their imputation model, MAR can never be proven (Schafer & Graham, 2002) and all inferences are always conditional on the assumption of MAR. This does not only hold for the computation of Bayes factor from data with missing values, this holds for all inferences using data with missing values.

It is important to note that the Bayes factor is a quantification of information in a data set and not an estimate of a population quantity. Therefore, the uncertainty in the Bayes factor computed from data with missing values consists of two parts: (a) uncertainty caused by the fact that the missing data mechanism is unknown which has been elaborated in this section; and (b) Monte Carlo error. The latter can be controlled by choosing an appropriate value for $Q$ as was elaborated at the end of the discussion of Example 1. Stated otherwise, there is no uncertainty due to sampling. Therefore, the variation in Bayes factors that can be observed in Figure 1 does not reflect uncertainty with respect to the value of the Bayes factor, it is a measure of the amount of information that is missing due to the presence of missing values: the larger the variation in Bayes factors computed for each imputed data matrix, the larger the amount of missing information.

## Example 4: Confirmatory Factor Analysis

The Holzinger and Swineford (1939) dataset consists of mental ability scores of 301 seventh- and eighth-grade children. A subset with nine variables is widely used in the literature. It contains scores on: visual perception $\boldsymbol{x}_1$; cubes $\boldsymbol{x}_2$; lozenges $\boldsymbol{x}_3$; paragraph comprehension $\boldsymbol{x}_4$; sentence completion $\boldsymbol{x}_5$; word meaning $\boldsymbol{x}_6$; speeded addition $\boldsymbol{x}_7$; speeded counting of dots $\boldsymbol{x}_8$; and speeded discrimination of straight and curved capitals $\boldsymbol{x}_9$. A common analysis model for these data is the following confirmatory factor model:

$$\begin{aligned} x_{pi} &= \alpha_p + \lambda_p t_{1i} + \epsilon_p \quad \text{for} \quad p = 1, 2, 3, \\ x_{pi} &= \alpha_p + \lambda_p t_{2i} + \epsilon_p \quad \text{for} \quad p = 4, 5, 6, \\ x_{pi} &= \alpha_p + \lambda_p t_{3i} + \epsilon_p \quad \text{for} \quad p = 7, 8, 9, \end{aligned} \tag{59}$$

where: $\boldsymbol{t}_1$ denotes a visual factor, $\boldsymbol{t}_2$ a textual factor, and $\boldsymbol{t}_3$ a speed factor with

$$\begin{pmatrix} t_{1i} \\ t_{2i} \\ t_{3i} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \gamma_{12} & \gamma_{13} \\ \gamma_{12} & 1 & \gamma_{23} \\ \gamma_{13} & \gamma_{23} & 1 \end{pmatrix} \right). \tag{60}$$

Note that, the $\boldsymbol{\gamma} = [\gamma_{12}, \gamma_{13}, \gamma_{23}]$ contains the correlations between the factors and $\boldsymbol{\omega}$ contains the intercepts $\alpha_p$, the factor loadings $\lambda_p$, and the residual variances from $\epsilon_p \sim \mathcal{N}(0, \sigma_p^2)$. The informative hypothesis of interest for this data set is

$$H_1 : \gamma_{12} > \gamma_{23}, \gamma_{13} > \gamma_{23}, \tag{61}$$

which expresses that the correlation between the textual and speed factor is the smallest of the three correlations. The Holzinger and Swineford (1939) data will be used to illustrate that the approach proposed in this article can also be used for other models than multiple regression. Missing data where created by replacing each observation by a missing with a probability of .20. Descriptive statistics are presented in Table 6.

The imputation model used was analogous to the one used for the multiple regression example displayed in Equation 3, that is, a multivariate normal distribution for the nine mental test scores. In the context of structural equation modeling and thus also the confirmatory factor analysis model displayed in Equations 59 and 60 this implies that the missing values are imputed under a saturated model, that is, a model that will perfectly reproduce the observed covariance matrix of the nine mental test scores. As is well known from the structural equation modeling literature (see, e.g., Bollen, 1989), such a model encompasses the analysis model, that is, the parameters of the analysis model are a function of the parameters of the imputation model. Furthermore, as is elaborated in Graham (2003) the analysis model can be extended into a "saturated correlations model" which is equivalent to the "saturated imputation model" to account for the presence of auxiliary variables. Stated otherwise, the parameters of the imputation model can be separated into the parameters of the analysis model plus extra parameters accounting for the presence of auxiliary variables, as is required for the approach presented in this article.

Using mice to obtain 1,000 imputed data matrices rendered $\hat{\gamma} = [39, .47, .26]$. Furthermore,

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \begin{pmatrix} .0056 & -.0003 & .0012 \\ -.0003 & .0230 & .0061 \\ .0012 & .0061 & .0065 \end{pmatrix}, \tag{62}$$

and the fraction of missing information is .42, that is, the effective sample size $N_0 = 301 - .42 \times 301$. Note that, about 20% missing values and a fraction of missing information of .42 may seem out of line. However, in the case of the classical correlation $\gamma_{xy}$ where the odd-numbered persons have a missing value for $y$ and the even numbered persons have a missing value for $x$, there are 50% missing values, while the fraction of information missing with respect to $\gamma_{xy}$ equals 1.0. Feeding this information into Bain

Table 6
*Descriptive Statistics for the Holzinger and Swineford (1939) Data*

| Variable | Mean | Minimum | Maximum | Number missing |
|---|---|---|---|---|
| $\boldsymbol{x}_1$ | 4.88 | .67 | 8.50 | 53 |
| $\boldsymbol{x}_2$ | 6.07 | 2.25 | 9.25 | 59 |
| $\boldsymbol{x}_3$ | 2.30 | .25 | 4.50 | 72 |
| $\boldsymbol{x}_4$ | 3.07 | .00 | 6.33 | 43 |
| $\boldsymbol{x}_5$ | 4.35 | 1.00 | 6.50 | 74 |
| $\boldsymbol{x}_6$ | 2.16 | .14 | 6.14 | 62 |
| $\boldsymbol{x}_7$ | 4.23 | 1.87 | 7.44 | 57 |
| $\boldsymbol{x}_8$ | 5.48 | 3.50 | 8.30 | 58 |
| $\boldsymbol{x}_9$ | 5.32 | 2.78 | 9.25 | 63 |

rendered $BF_{1u} = 3.58$, that is, there is some support in the data for $H_1$. If norm is used for imputation the resulting Bayes factor was almost identical, that is, $BF_{1u} = 3.58$. Note that the codes used to obtain the results presented in this section can be found in the online supplementary materials.

## Discussion

The computation of Bayes factors used for hypotheses testing from data with missing values has not previously received attention in the literature. In this article two specific Bayes factors have been presented. The adjusted fractional Bayes factor has been developed to evaluate informative hypotheses in the context of a wide variety of statistical models. It is implemented in the R package Bain (http://informative-hypotheses.sites.uu.nl/software/) which is currently being included in JASP (https://jasp-stats.org/). The R package BayesFactor (http://bayesfactorpcl.r-forge.r-project.org/) is a versatile tool for the evaluation of null-hypotheses and is also included in JASP. The results presented in this article extend the applicability of these Bayes factors for hypothesis evaluation, because empirical data often contain missing values. As is elaborated in Appendix A, the derivations presented in this article can also be applied to other Bayes factors. However, because these are not implemented in software packages, or, the software package does not render the necessary information, these cannot straightforwardly be used by psychological researchers.

It has to be noted that new options keep being added to R packages that can be used for multiple imputation. For example, recently, multiple imputation for two-level models has been added to mice Also both Bain and BayesFactor are packages that are actively maintained and to which new options keep being added. The interested reader is well-advised to monitor new developments in order to be up-to-date as to what is and isn't possible with these packages.

## References


Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association, 91,* 109–122. http://dx.doi.org/10.1080/01621459.1996.10476668

Berger, J. O., & Pericchi, L. R. (2004). Training samples in objective Bayesian model selection. *The Annals of Statistics, 32,* 841–869. http://dx.doi.org/10.1214/009053604000000229

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley-Interscience.

Chib, S. (1995). Marginal likelihood from Gibbs output. *Journal of the American Statistical Association, 90,* 1313–1321. http://dx.doi.org/10.1080/01621459.1995.10476635

Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association, 96,* 270–281. http://dx.doi.org/10.1198/016214501750332848

Cohen, J. (1994). The earth is round, $p < .05$. *American Psychologist, 49,* 997–1003. http://dx.doi.org/10.1037/0003-066X.49.12.997

Cumming, G. (2012). *Understanding the new statistics, effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Gilks, W. R. (1995). Discussion of fractional Bayes factors for model comparison (by O'Hagan). *Journal of the Royal Statistical Society Series B, 56,* 118–120.

Graham, J. W. (2003). Adding missing-data relevant variables to FIML based structural equation models. *Structural Equation Modeling, 10,* 80–100. http://dx.doi.org/10.1207/S15328007SEM1001_4

Gu, X. (2016). *Bayesian evaluation of informative hypotheses* (Doctoral dissertation). Retrieved from http://informative-hypotheses.sites.uu.nl/books-and-theses/

Gu, X., Hoijtink, H., & 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., & Hoijtink, H. (in press). Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. *British Journal of Mathematical and Statistical Psychology*.

Gu, X., Mulder, J., Dekovic, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods, 19,* 511–527. http://dx.doi.org/10.1037/met0000017

Hoijtink, H. (2012). *Informative hypotheses. Theory and practice for behavioral and social scientists*. Boca Raton, FL: Chapman and Hall/CRC.

Holzinger, K., & Swineford, F. (1939). *A study in factor analysis: The stability of a bifactor solution. Supplementary Educational Monograph, no. 48*. Chicago, IL: University of Chicago Press.

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., & Sterne, J. A. C. (2014). Joint modelling rationale for chained equations. *BMC Medical Research Methodology, 14,* 28. http://dx.doi.org/10.1186/1471-2288-14-28

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795. http://dx.doi.org/10.1080/01621459.1995.10476572

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality Constrained Analysis of Variance: A Bayesian approach. *Psychological Methods, 10,* 477–493. http://dx.doi.org/10.1037/1082-989X.10.4.477

Kolkman, M. E., Hoijtink, H., Kroesbergen, E. H., & Leseman, P. P. M. (2013). The role of executive functions in numerical magnitude skills. *Learning and Individual Differences, 24,* 145–151. http://dx.doi.org/10.1016/j.lindif.2013.01.004

Kropko, J., Goodrich, B., Gelman, A., & Hill, J. (2014). multiple imputation for continuous and categorical data: Comparing joint multivariate normal and conditional approaches. *Political Analysis, 22,* 497–519. http://dx.doi.org/10.1093/pan/mpu007

Liu, J., Gelman, A., Hill, J., Su, Y.-S., & Kropko, J. (2014). On the stationary distribution of iterative imputations. *Biometrika, 101,* 155–173. http://dx.doi.org/10.1093/biomet/ast044

Meng, X.-L. (1995). Multiple-imputation inferences with uncongenial source of input (with discussion). *Statistical Science, 10,* 538–573. http://dx.doi.org/10.1214/ss/1177010274

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods, 16,* 406–419. http://dx.doi.org/10.1037/a0024377

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

Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS a Fortran90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software, 46,* 2. http://dx.doi.org/10.18637/jss.v046.i02

Mulder, J., Hoijtink, H., & Klugkist, I. (2010). Equality and inequality constrained multivariate linear models: Objective model selection using constrained posterior priors. *Journal of Statistical Planning and Inference, 140,* 887–906. http://dx.doi.org/10.1016/j.jspi.2009.09.022

Mulder, J., & Wagenmakers, E.-J. (2016). Editors' introduction to the special issue "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments". *Journal of Mathematical Psychology, 72,* 1–5. http://dx.doi.org/10.1016/j.jmp.2016.01.002

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

Perez, J. M., & Berger, J. (2002). Expected posterior prior distributions for model selection. *Biometrika, 89,* 491–511. http://dx.doi.org/10.1093/biomet/89.3.491

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Royal, R. (1997). *Statistical evidence. A likelihood paradigm*. New York, NY: Chapman and Hall/CRC.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: Wiley and Sons.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman and Hall/CRC.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15. http://dx.doi.org/10.1177/096228029900800102

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. http://dx.doi.org/10.1037/1082-989X.7.2.147

Seaman, S. R., & Hughes, R. A. (2016). Relative efficiency of join-model and full-conditional-specification multiple imputation when conditional models are compatible: The general location model. *Statistical Methods in Medical Research*. Advance online publication. http://dx.doi.org/10.1177/0962280216665872

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah NJ: Erlbaum.

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specifications. *Statistical Methods in Medical Research, 16,* 219–242. http://dx.doi.org/10.1177/0962280206074463

Van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman and Hall/CRC.

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76,* 1049–1064. http://dx.doi.org/10.1080/10629360600810434

Van Schie, K., Van Veen, S. C., Engelhard, I. M., Klugkist, I., Hout van den, M. A. (2016). Blurring emotional memories using eye movements: Individual differences and speed of eye movements. *European Journal of Psychotraumatology*. Advance online publication. http://dx.doi.org/10.3402/ejpt.v7.29476

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 15,* 779–804. http://dx.doi.org/10.3758/BF03194105

Zhu, J., & Raghunathan, T. E. (2015). Covergence properties of a sequential regression multiple imputation algorithm. *Journal of the American Statistical Association, 110,* 1112–1124. http://dx.doi.org/10.1080/01621459.2014.948117

(*Appendices follow*)

# Appendix A

# Other Bayes Factors

## Fractional Bayes Factors

The fractional Bayes factor (O'Hagan, 1995) is different from the adjusted fractional Bayes factor in that the location of the fractional prior is not adjusted. It can be computed using Equations 17 and 21 (the latter without adjustment of the prior mean, that is, $\boldsymbol{\gamma}_{adj} = 0$) based on the following posterior and prior distributions:

$$g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m, \boldsymbol{X}_o) \propto \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \mid \boldsymbol{\theta})^{1-b(\boldsymbol{X}_m, \boldsymbol{X}_o)} \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \mid \boldsymbol{\theta})^{b(\boldsymbol{X}_m, \boldsymbol{X}_o)} h_u(\boldsymbol{\theta}), \tag{63}$$

and,

$$h_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m, \boldsymbol{X}_o) \propto \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \mid \boldsymbol{\theta})^{b(\boldsymbol{X}_m, \boldsymbol{X}_o)} h_u(\boldsymbol{\theta}), \tag{64}$$

with $h_u(\boldsymbol{\theta})$ a standard noninformative prior:

$$BF_{ku} \approx \frac{\frac{1}{Q}\sum_{q=1}^{Q} \int_{\boldsymbol{\theta} \in H_k} g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\theta}}{\frac{1}{Q}\sum_{q=1}^{Q} \int_{\boldsymbol{\theta} \in H_k} h_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\theta}} = \frac{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q}{\frac{1}{Q}\sum_{q=1}^{Q} c_k^q}, \tag{65}$$

that is, it is the ratio of the average of the fits $f_k^q$ and the average of the complexities $c_k^q$ computed with respect to $Q$ imputed data matrices. Because the prior mean is not adjusted, the fractional Bayes factor is only suited for the evaluation of informative hypotheses specified using equality constraints (Mulder, 2014). Because it is not implemented in a software package, it is not readily available.

## Bayes Factors Based on Expected Posterior Priors

Perez and Berger (2002) discuss Bayes factors based on expected posterior priors. They can be computed using Equations 17 and 21 (the latter without adjustment of the prior mean, that is, $\gamma_{adj} = 0$) based on the following posterior and prior distributions:

$$g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m, \boldsymbol{X}_o) \propto \ell(\boldsymbol{X}_m, \boldsymbol{X}_o \mid \boldsymbol{\theta}) h_u(\boldsymbol{\theta} \mid mts(\boldsymbol{X}_m, \boldsymbol{X}_o)), \tag{66}$$

where $h_u(\boldsymbol{\theta} \mid mts(\boldsymbol{X}_m, \boldsymbol{X}_o))$ denotes the expected posterior prior which is based on multiple minimal training samples (*mts*, Berger and Pericchi, 1996, 2004) obtained from $\boldsymbol{X}_m, \boldsymbol{X}_o$. In close analogy with the fractional Bayes factor where a fraction of the likelihood is used to specify the prior distribution, it follows that

$$BF_{ku} \approx \frac{\frac{1}{Q}\sum_{q=1}^{Q} \int_{\boldsymbol{\theta} \in H_k} g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\theta}}{\frac{1}{Q}\sum_{q=1}^{Q} \int_{\boldsymbol{\theta} \in H_k} h_u(\boldsymbol{\theta} \mid mts(\boldsymbol{X}_m^q, \boldsymbol{X}_o)) d\boldsymbol{\theta}} = \frac{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q}{\frac{1}{Q}\sum_{q=1}^{Q} c_k^q}, \tag{67}$$

that is, the ratio of the average of the fits $f_k^q$ and the average of the complexities $c_k^q$ computed with respect to $Q$ imputed data matrices. Mulder, Hoijtink, and Klugkist (2010) show that this Bayes factor is only suited for the evaluation of informative hypotheses specified using equality constraints. Because it is not implemented in a software package, it is not readily available.

Mulder et al. (2012) elaborate how the expected posterior prior can be adjusted such that it is suited for the evaluation of informative hypotheses specified using equality and inequality constraints. This prior is called the approximate adjusted expected posterior prior $h_u(w(\boldsymbol{\theta}) \mid mts(\boldsymbol{X}_m, \boldsymbol{X}_o))$, where $w(\cdot)$ denotes a transformation of $\boldsymbol{\theta}$ such that the prior distribution of $\boldsymbol{\theta}$ is normally distributed with mean $\boldsymbol{\theta}_B$ and restricted covariance matrix $\boldsymbol{\tau}_B$. This results in:

$$\begin{aligned} BF_{ku} &\approx \frac{\int_{\boldsymbol{\theta} \in H_m} \frac{1}{Q}\sum_{q=1}^{Q} g_u(\boldsymbol{\theta} \mid \boldsymbol{X}_m^q, \boldsymbol{X}_o) d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta} \in H_m} \frac{1}{Q}\sum_{q=1}^{Q} h_u(w(\boldsymbol{\theta}) \mid mts(\boldsymbol{X}_m^q, \boldsymbol{X}_o)) d\boldsymbol{\theta}} \\ &= \frac{\frac{1}{Q}\sum_{q=1}^{Q} f_k^q(w(\boldsymbol{\theta}))}{\frac{1}{Q}\sum_{q=1}^{Q} c_k^q(w(\boldsymbol{\theta}))}, \end{aligned} \tag{68}$$

where the notation stresses that the fit $f_k^q(w(\boldsymbol{\theta}))$ and the complexity $c_k^q(w(\boldsymbol{\theta}))$ are computed using the $q$-th imputed data matrix and a $w(\boldsymbol{\theta})$ that is based on the information in $\boldsymbol{X}_o$ and therefore the same for each imputed data matrix.

The approximate adjusted expected posterior prior is implemented in the software package BIEMS (http://informative-hypotheses.sites.uu.nl/software/). As is shown in Equation 68, what is needed to compute the corresponding Bayes factor from data with missing values are the fit and complexity based on a $w(\boldsymbol{\theta})$ that is the same for each of the $q = 1, \ldots, Q$ imputed data matrices. BIEMS only renders the fit and complexity (and only for inequality constrained hypotheses) based on $w^q(\boldsymbol{\theta})$, that is, the transformation is determined using the $q$-th imputed data matrix and not using only the observed values. Therefore, BIEMS cannot be used to compute Equation 68. Researchers currently using BIEMS for the evaluation of informative hypotheses are advised to change to Bain or BayesFactor if their data are incomplete.

(*Appendices continue*)

## Appendix B

## Derivation of the Bayes Factor for Hypothesis Evaluation

Chib's (1995) representation of the Bayes factor comparing the informative hypothesis $H_k$ with the unconstrained hypothesis $H_u$ can be written as:

$$BF_{ku} = \frac{p(\boldsymbol{X}_o, \boldsymbol{Z}_o \mid \boldsymbol{\theta}, \boldsymbol{\eta}) h_k(\gamma + \gamma_{adj}, \omega, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)}{g_k(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)} \Big/ \frac{p(\boldsymbol{X}_o, \boldsymbol{Z}_o \mid \boldsymbol{\theta}, \boldsymbol{\eta}) h_u(\gamma + \gamma_{adj}, \boldsymbol{\omega}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)}{g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o)}, \tag{69}$$

where, $h_k(\cdot)$ and $h_u(\cdot)$ denote the prior distribution of the parameters for the informative and unconstrained hypotheses, respectively, and $g_k(\cdot)$ and $g_u(\cdot)$ denote the posterior distribution for the informative and unconstrained hypotheses, respectively.

As is highlighted in Equation 4, $H_k$ for $k = 1, \ldots, K$ are hypotheses obtained by imposing equality and/or inequality constraints on a parameter vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_J]$. Each of these hypotheses is nested with $H_u : \gamma_1, \ldots, \gamma_J$, that is, a hypothesis without constraints on the parameters.

The prior distributions $h_k(\cdot)$ and $g_k(\cdot)$ are those parts of $h_u(\cdot)$ and $g_u(\cdot)$, respectively, in agreement with $H_k$ (Hoijtink, 2012, p. 59), that is,

$$\begin{aligned} h_k(\gamma + \gamma_{adj}, \omega, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) &= \frac{h_u(\gamma + \gamma_{adj}, \omega, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) I(\gamma \in H_k)}{\int_{\gamma,\omega,\eta} h_u(\gamma + \gamma_{adj}, \omega, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) I(\gamma \in H_k) d\gamma d\omega d\eta} \\ &= \frac{h_u(\gamma + \gamma_{adj}, \omega, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) I(\gamma \in H_k)}{c_k}, \end{aligned} \tag{70}$$

where $I(\cdot)$ equals one is the argument is true and zero otherwise, and,

$$\begin{aligned} g_k(\boldsymbol{\theta}, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) &= \frac{g_u(\boldsymbol{\theta}, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) I(\boldsymbol{\theta} \in H_k)}{\int_{\boldsymbol{\theta},\eta} g_u(\boldsymbol{\theta}, \eta \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) I(\boldsymbol{\theta} \in H_k) d\boldsymbol{\theta} d\boldsymbol{\eta}} \\ &= \frac{g_u(\boldsymbol{\theta}, \boldsymbol{\eta} \mid \boldsymbol{X}_o, \boldsymbol{Z}_o) I(\boldsymbol{\theta} \in H_k)}{f_k}. \end{aligned} \tag{71}$$

Substitution of $h_k(\cdot)$ and $g_k(\cdot)$ in Equation 69 with Equation 70 and 71, respectively, renders

$$BF_{ku} = f_k / c_k. \tag{72}$$

Received April 4, 2017
Revision received January 9, 2018
Accepted February 12, 2018 ■