Journal of Mathematical Psychology 72 (2016) 158–170

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Automatic Bayes factors for testing variances of two independent normal distributions

Florian Böing-Messing *, Joris Mulder

Department of Methodology and Statistics, Tilburg University, Postbus 90153, 5000 LE Tilburg, The Netherlands

HIGHLIGHTS

- We develop three automatic Bayes factors for testing two variances.
- We consider a fractional, a balanced, and an adjusted fractional Bayes approach.
- The Bayes factors do not require prior elicitation and are thus fully automatic.
- We evaluate the methods based on theoretical properties and numerical performance.
- The adjusted fractional Bayes factor performs best overall.

ARTICLE INFO

Article history:
Available online 1 September 2015

Keywords:
Automatic Bayes factor
Automatic prior
Balancedness
Fractional Bayes factor
Homogeneity of variance
Inequality constrained hypothesis
Occam's razor

ABSTRACT

Researchers are frequently interested in testing variances of two independent populations. We often would like to know whether the population variances are equal, whether population 1 has smaller variance than population 2, or whether population 1 has larger variance than population 2. In this article we consider the Bayes factor, a Bayesian model selection and hypothesis testing criterion, for this multiple hypothesis test. Application of Bayes factors requires specification of prior distributions for the model parameters. Automatic Bayes factors circumvent the difficult task of prior elicitation by using data-driven mechanisms to specify priors in an automatic fashion. In this article we develop different automatic Bayes factors for testing two variances: first we apply the fractional Bayes factor (FBF) to the testing problem. It is shown that the FBF does not always function as Occam's razor. Second we develop a new automatic balanced Bayes factor with equal priors for the variances. Third we propose a Bayes factor based on an adjustment of the marginal likelihood in the FBF approach. The latter two methods always function as Occam's razor. Through theoretical considerations and numerical simulations it is shown that the third approach provides strongest evidence in favor of the true hypothesis.

1. Introduction

Researchers are frequently interested in comparing two independent populations on a continuous outcome measure. Traditionally, the focus has been on comparing means, whereas variances are mostly considered nuisance parameters. However, by regarding variances as mere nuisance parameters, one runs the risk of overlooking important information in the data. The variability of a population is a key characteristic which can be the core of a research question. For example, psychological research frequently investigates differences in variability between males and females (e.g. Arden & Plomin, 2006; Borkenau, Hřebíčková, Kuppens, Realo, & Allik, 2013; Feingold, 1992).

In this article we consider a Bayesian hypothesis test on the variances of two independent populations. The Bayes factor is a well-known Bayesian criterion for model selection and hypothesis testing (Jeffreys, 1961; Kass & Raftery, 1995). Unlike the *p*-value, which is often misinterpreted as an error probability (Hubbard & Armstrong, 2006), the Bayes factor has a straightforward interpretation as the relative evidence in the data in favor of a hypothesis as compared to another hypothesis. Moreover, contrary to *p*-values, the Bayes factor is able to quantify evidence in favor of a null hypothesis (Wagenmakers, 2007). Another useful property, which is not shared by *p*-values, is that the Bayes factor can straightforwardly be used for testing multiple hypotheses simultaneously (Berger & Mortera, 1999). These and other notions have resulted in a considerable development of Bayes factors for frequently

* Corresponding author.
E-mail address: florian.boeingmessing@gmail.com (F. Böing-Messing).

http://dx.doi.org/10.1016/j.jmp.2015.08.001
0022-2496/

encountered testing problems in the last decade. For example, Klugkist, Laudy, and Hoijtink (2005) proposed Bayes factors for testing analysis of variance models. Rouder, Speckman, Sun, Morey, and Iverson (2009) proposed a Bayesian t-test. Mulder, Hoijtink, and de Leeuw (2012) developed a software program for Bayesian testing of (in)equality constraints on means and regression coefficients in the multivariate normal linear model, and Wetzels and Wagenmakers (2012) proposed Bayesian tests for correlation coefficients. The goal of this article is to extend this literature by developing Bayes factors for testing variances. For more interesting references we also refer the reader to the special issue 'Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments' in the *Journal of Mathematical Psychology* in which this article appeared (Mulder & Wagenmakers, in preparation).

In applying Bayes factors for hypothesis testing, we need to specify a prior distribution of the model parameters under every hypothesis to be tested. A prior distribution is a probability distribution describing the probability of the possible parameter values before observing the data. In the case of testing two variances, we need to specify a prior for the common variance under the null hypothesis and for the two unique variances under the alternative hypothesis. Specifying priors is a difficult task from a practical point of view, and it is complicated by the fact that we cannot use noninformative improper priors for parameters to be tested because the Bayes factor would then be undefined (Jeffreys, 1961). This has stimulated researchers to develop Bayes factors which do not require prior elicitation using external prior information. Instead, these so-called automatic Bayes factors use information from the sample data to specify priors in an automatic fashion. So far, however, no automatic Bayes factors have been developed for testing variances.

In this article we develop three types of automatic Bayes factors for testing variances of two independent normal populations. We first consider the fractional Bayes factor (FBF) of O'Hagan (1995) and apply it for the first time to the problem of testing variances. In the FBF methodology the likelihood of the complete data is divided into two fractions: one for specifying the prior and one for testing the hypotheses. However, it has been shown (e.g. Mulder, 2014b) that the FBF may not be suitable for testing inequality constrained hypotheses (e.g. variance 1 is smaller than variance 2) because it may not function as Occam's razor. In other words, the FBF may not prefer the simpler hypothesis when two hypotheses fit the data equally well. This is a consequence of the fact that in the FBF the automatic prior is located at the likelihood of the data. We develop two novel solutions to this problem: the first is an automatic Bayes factor with equal automatic priors for both variances under the alternative hypothesis. This methodology is related to the constrained posterior priors approach of Mulder, Hoijtink, and Klugkist (2010). The second novel solution is an automatic Bayes factor based on adjusting the definition of the FBF such that the resulting automatic Bayes factor always functions as Occam's razor. This approach is related to the work of Mulder (2014b), with the difference that our method results in stronger evidence in favor of a true null hypothesis.

The remainder of this article is structured as follows. In the next section we provide details on the normal model to be used and introduce the hypotheses we shall be concerned with. We then discuss five theoretical properties which are used for evaluating the automatic Bayes factors. Following this, we develop the three automatic Bayes factors and evaluate them according to the theoretical properties. Subsequently, the performance of the Bayes factors is investigated by means of a small simulation study. We conclude the article with an application of the Bayes factors to two empirical data examples and a discussion of possible extensions and limitations of our approaches.

2. Model and hypotheses

We assume that the outcome variable of interest, X, is normally distributed in both populations:

$$X_j \sim \mathrm{N}\left(\mu_j, \sigma_j^2\right), \quad j = 1, 2, \tag{1}$$

where j is the population index and μ_j and σ_j^2 are the population-specific parameters. The unknown parameter in this model is $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)' = \left(\left(\mu_1, \mu_2,\right)', \left(\sigma_1^2, \sigma_2^2\right)'\right)' \in \mathbb{R}^2 \times \Omega_u$, where $\Omega_u := \left(\mathbb{R}^+\right)^2$ is the unconstrained parameter space of $\boldsymbol{\sigma}^2$.

In this article we shall be concerned with testing the following nonnested (in)equality constrained hypotheses against one another:

$$\begin{array}{ll} H_0\colon \sigma_1^2 = \sigma_2^2 = \sigma^2, & H_0\colon \sigma^2 \in \Omega_0 := \mathbb{R}^+, \\ H_1\colon \sigma_1^2 < \sigma_2^2, & \Leftrightarrow H_1\colon \boldsymbol{\sigma}^2 \in \Omega_1 := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\right\}, \\ H_2\colon \sigma_1^2 > \sigma_2^2, & H_2\colon \boldsymbol{\sigma}^2 \in \Omega_2 := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > \sigma_2^2\right\}, \end{array} \tag{2}$$

where $\Omega_1, \Omega_2 \subset \Omega_u$ and Ω_0 denote the parameter spaces under the corresponding (in)equality constrained hypotheses.

We made two choices in formulating the hypotheses in Eq. (2). First, we do not test any constraints on the mean parameters μ_1 and μ_2. This is because the objective of this article is to provide a Bayesian alternative to the classical frequentist procedures for testing two variances. For a general framework for testing (in)equality constrained hypotheses on mean parameters, see, for example, Mulder et al. (2012). The second choice we made is to divide the classical alternative hypothesis $H_a\colon \sigma_1^2 \neq \sigma_2^2 \Leftrightarrow H_a\colon \sigma_1^2 < \sigma_2^2 \vee \sigma_1^2 > \sigma_2^2$ into two separate hypotheses, $H_1\colon \sigma_1^2 < \sigma_2^2$ and $H_2\colon \sigma_1^2 > \sigma_2^2$ ($\vee$ denotes logical disjunction and reads "or"). The advantage of this approach is that it allows us to quantify and compare the evidence in favor of a negative effect (H_1) and a positive effect (H_2). This is of great interest to applied researchers, who would often like to know not only whether there is an effect, but also in what direction.

Another hypothesis we will consider is the unconstrained hypothesis

$$H_u\colon \sigma_1^2, \sigma_2^2 > 0 \Leftrightarrow H_u\colon \boldsymbol{\sigma}^2 \in \Omega_u = \left(\mathbb{R}^+\right)^2. \tag{3}$$

This hypothesis is not of substantial interest to us because it is entirely covered by the hypotheses in Eq. (2). In other words, $\{H_0, H_1, H_2\}$ is a partition of H_u. The unconstrained hypothesis will be used to evaluate theoretical properties of the priors and Bayes factors such as balancedness and Occam's razor (discussed in the next section).

3. Properties for the automatic priors and Bayes factors

Based on the existing literature on automatic Bayes factors, we shall focus on the following theoretical properties when evaluating the automatic priors and Bayes factors:

1. *Proper priors: The priors must be proper probability distributions*. When using improper priors on parameters that are tested, the resulting Bayes factors depend on unspecified constants (see, for instance, O'Hagan, 1995). Improper priors may only be used on common nuisance parameters that are present under all hypotheses to be tested (Jeffreys, 1961).
2. *Minimal information: Priors under composite hypotheses should contain the information of a minimal study*. Using arbitrarily vague priors gives rise to the Jeffreys–Lindley paradox (Jeffreys, 1961; Lindley, 1957), whereas priors containing too much information about the parameters will dominate the data. Therefore it is often suggested to let the prior contain the

information of a minimal study (e.g. Berger & Pericchi, 1996; O'Hagan, 1995; Spiegelhalter & Smith, 1982). A minimal study is the smallest possible study (in terms of sample size) for which all free parameters under all hypotheses are identifiable. If prior information is absent (as is usually the case when automatic Bayes factors are considered), then a prior containing minimal information is a reasonable starting point.

3. *Scale invariance: The Bayes factors should be invariant under rescaling of the data*. In other words, the Bayes factors should not depend on the scale of the outcome variable. This is important because when comparing, say, the heterogeneity of ability scores of males and females, it should not matter if the ability test has a scale from 0 to 10 or from 0 to 100.
4. *Balancedness: The prior under the unconstrained hypothesis should be balanced*. If we denote $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$, then the unconstrained hypothesis can be written as $H_u\colon \eta \in \mathbb{R}$. The prior for η under H_u should be symmetric about 0 and nonincreasing in $|\eta|$ (e.g. Berger & Delampady, 1987). Following Jeffreys (1961), we shall refer to a prior satisfying these properties as a balanced prior. A balanced prior can be considered objective in two respects: first, the symmetry ensures that neither a positive nor a negative effect is preferred a priori. Second, the nonincreasingness ensures that no other values but 0 are treated as special.
5. *Occam's razor: The Bayes factors should function as Occam's razor*. Occam's razor is the principle that if two hypotheses fit the data equally well, then the simpler (i.e. less complex) hypothesis should be preferred. The principle is based on the empirical observation that simple hypotheses that fit the data are more likely to be correct than complicated ones. When testing nested hypotheses, Bayes factors automatically function as Occam's razor by balancing fit and complexity of the hypotheses (Kass & Raftery, 1995). When testing inequality constrained hypotheses, however, the Bayes factor does not always function as Occam's razor (Mulder, 2014a).

4. Automatic Bayes factors

The Bayes factor is a Bayesian hypothesis testing criterion that is related to the likelihood ratio statistic. It is equal to the ratio of the marginal likelihoods under two competing hypotheses:

$$B_{pq} = \frac{m_p\left(\mathbf{x}\right)}{m_q\left(\mathbf{x}\right)}, \tag{4}$$

where B_{pq} denotes the Bayes factor comparing hypotheses H_p and H_q, and $m_p\left(\mathbf{x}\right)$ is the marginal likelihood under hypothesis H_p as a function of the data $\mathbf{x}$.

4.1. Fractional Bayes factor

The fractional Bayes factor introduced by O'Hagan (1995) is a general, automatic method for comparing two statistical models or hypotheses. In this article we apply it for the first time to the problem of testing variances. We use the superscript F to refer to the FBF.

4.1.1. Marginal likelihoods

The FBF marginal likelihood under hypothesis $H_p,\ p = 0, 1, 2, u$, is given by

$$m_p^F\left(\mathbf{b}, \mathbf{x}\right) = \frac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)\pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}{\int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}\pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2}, \tag{5}$$

where $p = u$ refers to the unconstrained hypothesis (with a slight abuse of notation), and under H_0 the variance parameter $\boldsymbol{\sigma}^2$ is a scalar containing only the common variance σ^2. Here $\pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ is the noninformative Jeffreys prior on $\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)'$. Under H_0 it is $\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right) \propto \sigma^{-2}$, while under H_u we have $\pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) \propto \sigma_1^{-2}\sigma_2^{-2}$. Under $H_p,\ p = 1, 2$, the Jeffreys prior is $\pi_p^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) \propto \sigma_1^{-2}\sigma_2^{-2}1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right)$, where $1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right)$ is the indicator function which is 1 if $\boldsymbol{\sigma}^2 \in \Omega_p$ and 0 otherwise. The expression $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}$ denotes a fraction of the likelihood, the cornerstone of the FBF methodology. Let $\mathbf{x}_j = \left(x_{1j}, \ldots, x_{n_jj}\right)'$ be a vector of n_j observations coming from X_j. Fractions of the likelihoods under the four hypotheses are given by

$$\begin{aligned}
f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}} &:= f\left(\mathbf{x}_1|\mu_1, \sigma^2\right)^{b_1} f\left(\mathbf{x}_2|\mu_2, \sigma^2\right)^{b_2}, \\
f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} &:= f\left(\mathbf{x}_1|\mu_1, \sigma_1^2\right)^{b_1} f\left(\mathbf{x}_2|\mu_2, \sigma_2^2\right)^{b_2}, \\
f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} &:= f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}} 1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right), \quad p = 1, 2,
\end{aligned} \tag{6}$$

where

$$f\left(\mathbf{x}_j|\mu_j, \sigma_j^2\right)^{b_j} = \left(\prod_{i=1}^{n_j} \mathrm{N}\left(x_{ij}|\mu_j, \sigma_j^2\right)\right)^{b_j} \tag{7}$$

is a fraction of the likelihood of population j (e.g. Berger & Pericchi, 2001). Here $b_1 \in (1/n_1, 1]$ and $b_2 \in (1/n_2, 1]$ are population-specific proportions to be determined by the user, and by using $\mathbf{b} = (b_1, b_2)'$ as a superscript we slightly abuse notation. We obtain the full likelihood $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ by setting $b_1 = b_2 = 1$.

Plugging $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)$, $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$, and $\pi_0^N\left(\boldsymbol{\mu}, \sigma^2\right)$ into Eq. (5), we obtain the marginal likelihood under H_0 after some algebra (see Appendix A) as

$$\begin{aligned}
&m_0^F\left(\mathbf{b}, \mathbf{x}\right) \\
&= \frac{\left(b_1b_2\right)^{\frac{1}{2}}\Gamma\left(\frac{n_1+n_2-2}{2}\right)\left(b_1\left(n_1-1\right)s_1^2 + b_2\left(n_2-1\right)s_2^2\right)^{\frac{b_1n_1+b_2n_2-2}{2}}}{\pi^{\frac{n_1(1-b_1)+n_2(1-b_2)}{2}}\Gamma\left(\frac{b_1n_1+b_2n_2-2}{2}\right)\left(\left(n_1-1\right)s_1^2 + \left(n_2-1\right)s_2^2\right)^{\frac{n_1+n_2-2}{2}}},
\end{aligned} \tag{8}$$

where Γ denotes the gamma function, and $s_j^2 = \frac{1}{n_j-1}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2$ is the sample variance of $\mathbf{x}_j,\ j = 1, 2$. The marginal likelihoods under H_1 and H_2 are functions of the marginal likelihood under H_u, which is given by

$$m_u^F\left(\mathbf{b}, \mathbf{x}\right) = \frac{\pi^{-\frac{n_1(1-b_1)+n_2(1-b_2)}{2}} b_1^{\frac{b_1n_1}{2}} b_2^{\frac{b_2n_2}{2}} \Gamma\left(\frac{n_1-1}{2}\right)\Gamma\left(\frac{n_2-1}{2}\right)}{\Gamma\left(\frac{b_1n_1-1}{2}\right)\Gamma\left(\frac{b_2n_2-1}{2}\right)\left(\left(n_1-1\right)s_1^2\right)^{\frac{n_1(1-b_1)}{2}}\left(\left(n_2-1\right)s_2^2\right)^{\frac{n_2(1-b_2)}{2}}}. \tag{9}$$

For the marginal likelihoods under H_1 and H_2 we then have

$$m_p^F\left(\mathbf{b}, \mathbf{x}\right) = \frac{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)}{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)} m_u^F\left(\mathbf{b}, \mathbf{x}\right), \quad p = 1, 2. \tag{10}$$

Here $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)$ and $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)$ denote the probability that $\boldsymbol{\sigma}^2$ is in Ω_p given the complete data $\mathbf{x}$ or a fraction thereof (for which we use the notation $\mathbf{x}^{\mathbf{b}}$). The exact expressions for the two probabilities are given in Eqs. (B.1) and (B.2) in Appendix B. The derivation of Eqs. (9) and (10) is analogous to that of Eq. (8) given in Appendix A.

4.1.2. Evaluation of the method

We will now evaluate the FBF according to the five properties discussed in Section 3:

1. *Proper priors*. First, note that the marginal likelihood in Eq. (5) can be rewritten as

$$m_p^F(\mathbf{b},\mathbf{x}) = \int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}} \times \frac{f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)^{\mathbf{b}}\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)}{\int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)^{\mathbf{b}}\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2} d\boldsymbol{\mu}d\boldsymbol{\sigma}^2 = \int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}}\pi_p^F\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2, \quad (11)$$

where we use the superscript $\mathbf{1}-\mathbf{b} = (1-b_1, 1-b_2)'$ analogously to $\mathbf{b}$ in Eq. (6). Here $\pi_p^F\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) \propto f_p\left(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)^{\mathbf{b}}\pi_p^N\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)$ is a posterior prior obtained by updating the Jeffreys prior with a fraction of the likelihood. It can be considered the automatic prior implied by the FBF approach and is proper if $b_1n_1 + b_2n_2 > 2$ under H_0 and $b_jn_j > 1$, $j = 1, 2$, under H_1, H_2, and H_u. We use the notation $\mathbf{x}^{\mathbf{b}}$ to indicate that it is based on a fraction $\mathbf{b}$ of the likelihood of the complete sample data $\mathbf{x}$.
2. *Minimal information*. A minimal study consists of four observations, two from each population. This is because we need two observations from population j for $\left(\mu_j, \sigma_j^2\right)'$ to be identifiable. We can make the priors contain the information of a minimal study by setting $\mathbf{b} = (2/n_1, 2/n_2)'$ (O'Hagan, 1995).
3. *Scale invariance*. Multiplying all observations in $\mathbf{x}_j$ by a constant w results in a sample variance of $w^2s_j^2$, $j = 1, 2$. Plugging $w^2s_j^2$ into the formulas for the marginal likelihoods in Eqs. (8) and (9) does not change the resulting Bayes factors. Thus the FBF is scale invariant.
4. *Balancedness*. The marginal unconstrained prior on $\boldsymbol{\sigma}^2$ implied by the FBF approach is given by

$$\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|\nu_1,\tau_1^2\right)\text{Inv-}\chi^2\left(\sigma_2^2|\nu_2,\tau_2^2\right), \quad (12)$$

where

$$\nu_j = b_jn_j - 1 \quad \text{and} \quad \tau_j^2 = \frac{b_j\left(n_j-1\right)s_j^2}{b_jn_j - 1}, \quad j = 1, 2. \quad (13)$$

Here $\text{Inv-}\chi^2\left(\nu,\tau^2\right)$ is the scaled inverse-χ^2 distribution with degrees of freedom hyperparameter $\nu > 0$ and scale hyperparameter $\tau^2 > 0$ (Gelman, Carlin, Stern, & Rubin, 2004). The corresponding unconstrained prior on $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$, $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$, is balanced if and only if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$ ($\wedge$ denotes logical conjunction and reads "and"; see Appendix C for a proof). In practice the sample sizes and sample variances will commonly be such that $\neg\left(\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2\right)$, which is why $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$ will commonly be unbalanced ($\neg$ denotes logical negation and reads "not"). Fig. 1 illustrates this. The figure shows the priors on $\boldsymbol{\sigma}^2$ (top row) and η (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. It can be seen that $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$ is only balanced if $s_2^2 = s_1^2 = 1$, in which case $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$. For $s_2^2 \in \{4, 16\}$ it is shifted to the left (i.e. it is not skewed).
5. *Occam's razor*. The Bayes factors B_{1u}^F and B_{2u}^F should function as Occam's razor by favoring the simplest hypothesis that is in line with the data. This, however, is not the case, as Fig. 2 illustrates. The plot shows B_{1u}^F (solid line) and B_{2u}^F (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. It can be seen that B_{1u}^F and B_{2u}^F approach 1 for very large and very small s_2^2, respectively. Thus B_{1u}^F and B_{2u}^F are indecisive despite the data strongly supporting the more parsimonious inequality constrained hypothesis. This undesirable property is a direct consequence of the fact that the unconstrained prior is located at the likelihood of the data.

4.2. Balanced Bayes factor

In the previous section we have seen that the FBF involves two problems: the marginal unconstrained prior $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ is unbalanced and the Bayes factors B_{pu}^F and B_{p0}^F, $p = 1, 2$, do not function as Occam's razor. In this section we propose a solution to these problems which we refer to as the balanced Bayes factor (BBF). The BBF is a new automatic Bayes factor for testing variances of two independent normal distributions that satisfies all five properties discussed in Section 3. The BBF approach is related to the constrained posterior priors approach of Mulder et al. (2010) with the exception that the latter uses empirical training samples for prior specification instead of a fraction of the likelihood. The fractional approach of the BBF is therefore computationally less demanding. We use the superscript B to refer to the BBF.

4.2.1. Marginal likelihoods

In the FBF approach the marginal unconstrained prior $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|\nu_1,\tau_1^2\right)\text{Inv-}\chi^2\left(\sigma_2^2|\nu_2,\tau_2^2\right)$ is balanced if and only if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$, which in practice will rarely be the case. The main idea of the BBF thus is to replace $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ with a marginal unconstrained prior $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|\nu,\tau^2\right)\text{Inv-}\chi^2\left(\sigma_2^2|\nu,\tau^2\right)$ with common hyperparameters ν and τ^2. This way $\pi_u^B\left(\eta|\mathbf{x}^{\mathbf{b}}\right)$ is balanced by definition (see Appendix C). As with the FBF, we shall use information from the sample data $\mathbf{x}$ to define ν and τ^2: first we assume that $\sigma_1^2 = \sigma_2^2$ and update the Jeffreys prior with a fraction of the likelihood under H_0, $f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)^{\mathbf{b}}$. Note that this results in the FBF posterior prior $\pi_0^F\left(\boldsymbol{\mu},\sigma^2|\mathbf{x}^{\mathbf{b}}\right)$. Next, we obtain the marginal posterior prior on σ^2 by integrating out $\boldsymbol{\mu}$:

$$\pi_0^F\left(\sigma^2|\mathbf{x}^{\mathbf{b}}\right) = \int_{\mathbb{R}^2}\pi_0^F\left(\boldsymbol{\mu},\sigma^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu} = \text{Inv-}\chi^2\left(\sigma^2|\nu_\bullet,\tau_\bullet^2\right), \quad (14)$$

where

$$\nu_\bullet = b_1n_1 + b_2n_2 - 2 \quad \text{and} \quad \tau_\bullet^2 = \frac{b_1\left(n_1-1\right)s_1^2 + b_2\left(n_2-1\right)s_2^2}{b_1n_1 + b_2n_2 - 2}. \quad (15)$$

We use the subscript $\bullet$ to indicate that the hyperparameters $\nu_\bullet$ and $\tau_\bullet^2$ combine information from both samples $\mathbf{x}_1$ and $\mathbf{x}_2$. We propose using the distribution in Eq. (14) as the prior on both σ_1^2 and σ_2^2 under H_u, giving us the BBF marginal unconstrained prior on $\boldsymbol{\sigma}^2$ as

$$\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \pi_0^F\left(\sigma_1^2|\mathbf{x}^{\mathbf{b}}\right)\pi_0^F\left(\sigma_2^2|\mathbf{x}^{\mathbf{b}}\right), \quad (16)$$

with $\pi_0^F\left(\sigma_j^2|\mathbf{x}^{\mathbf{b}}\right)$ as in Eq. (14). Note that b_1 and b_2 need to be specified such that $b_1n_1 + b_2n_2 > 2$ for $\nu_\bullet$ to be positive. With the marginal unconstrained prior at hand, we define the joint prior on $\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)'$ under H_u as

$$\pi_u^B\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)\pi^N\left(\boldsymbol{\mu}\right), \quad (17)$$

with $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ as in Eq. (16). Here $\pi^N\left(\boldsymbol{\mu}\right) \propto 1$ is the Jeffreys prior for $\boldsymbol{\mu}$, which we may use since in our testing problem $\boldsymbol{\mu}$ is a common nuisance parameter that is present under all hypotheses. We shall define the BBF priors under H_1 and H_2 as truncations of the prior under H_u (Berger & Mortera, 1999; Klugkist et al., 2005):

$$\pi_p^B\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right) = \frac{1}{P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)}\pi_u^B\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right) = 2\cdot\pi_u^B\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)1_{\Omega_p}\left(\boldsymbol{\sigma}^2\right), \quad p = 1, 2, \quad (18)$$

where

$$P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right) = \int_{\Omega_p}\int_{\mathbb{R}^2}\pi_u^B\left(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2 = \int_{\Omega_p}\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\sigma}^2 = 0.5. \quad (19)$$

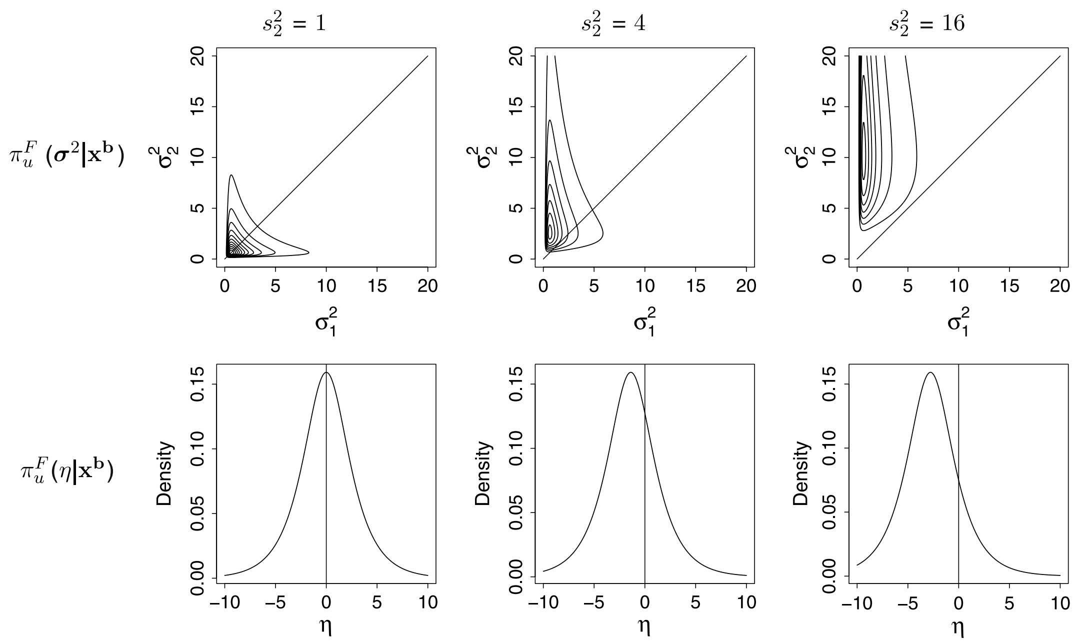


Fig. 1. The marginal unconstrained FBF prior $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ (top row) and the corresponding prior $\pi_u^F\left(\eta = \log\left(\sigma_1^2/\sigma_2^2\right)|\mathbf{x}^{\mathbf{b}}\right)$ (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. The prior $\pi_u^F(\eta|\mathbf{x}^{\mathbf{b}})$ is only balanced when $s_2^2 = s_1^2 = 1$.

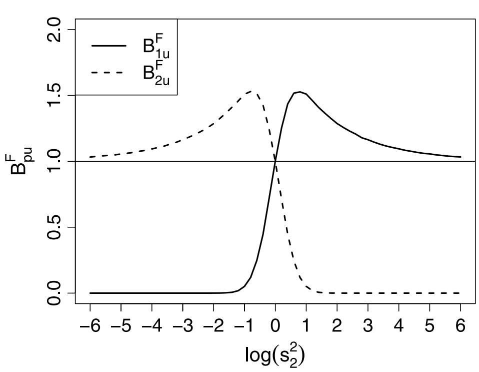


Fig. 2. Bayes factors B_{1u}^F (solid line) and B_{2u}^F (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. The Bayes factors approach 1 for very large and very small s_2^2, respectively. That is, they do not favor the more parsimonious inequality constrained hypothesis even though it is strongly supported by the data. This shows that B_{1u}^F and B_{2u}^F do not function as Occam's razor.

We have $P^B\left(\boldsymbol{\sigma}^2 \in \Omega_1|\mathbf{x}^{\mathbf{b}}\right) = P^B\left(\boldsymbol{\sigma}^2 \in \Omega_2|\mathbf{x}^{\mathbf{b}}\right) = 0.5$ because $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$ is the product of two identical scaled inverse-χ^2 distributions. In Eq. (18) the inverse $1/P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^{\mathbf{b}}\right)$ acts as a normalizing constant. Eventually, we define the BBF prior under H_0 such that it is in line with the priors under H_1 and H_2:

$$\pi_0^B\left(\boldsymbol{\mu}, \sigma^2|\mathbf{x}^{\mathbf{b}}\right) = \pi_0^F\left(\sigma^2|\mathbf{x}^{\mathbf{b}}\right)\pi^N(\boldsymbol{\mu}), \tag{20}$$

with $\pi_0^F\left(\sigma^2|\mathbf{x}^{\mathbf{b}}\right)$ as in Eq. (14).

With the priors at hand we can now determine the marginal likelihoods. The BBF marginal likelihood under hypothesis $H_p,\ p = 0, 1, 2, u$, is given by

$$m_p^B(\mathbf{b}, \mathbf{x}) = \int_{\Omega_p}\int_{\mathbb{R}^2} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)\pi_p^B\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2. \tag{21}$$

Besides the prior, this formulation differs from the FBF marginal likelihood in another important aspect: in Eq. (11) we have seen that to compute the FBF marginal likelihood we implicitly factor the full likelihood as $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) = f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}} f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}$. Then a proper posterior prior is obtained using $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{b}}$, and the marginal likelihood is computed using the remaining fraction $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}}$. From Eq. (21) it can be seen that to compute the BBF marginal likelihoods we use the full likelihood $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ instead of $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^{\mathbf{1}-\mathbf{b}}$. That is, we first use $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$ to obtain the proper prior $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^{\mathbf{b}}\right)$, and subsequently we use $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ to compute the marginal likelihoods. This implies that we use the data twice, once for prior specification and once for hypothesis testing. We choose to do so for the following reason: we use the information in $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$ to specify the variance of the balanced prior, but not its location. This means that we use less information for prior specification than is actually contained in $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)^{\mathbf{b}}$. Therefore, the full likelihood $f_p\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)$ is used for hypothesis testing. The latter illustrates that the BBF approach differs fundamentally from standard automatic procedures such as the FBF in which the likelihood is explicitly divided into a training part and a testing part. This is reflected in the function of **b** in the FBF and the BBF: while in the FBF the **b** determines how the likelihood is divided, in the BBF it determines how much of the information in the data we want to use twice.

Now, plugging $f_0\left(\mathbf{x}|\boldsymbol{\mu}, \sigma^2\right)$ and $\pi_0^B\left(\boldsymbol{\mu}, \sigma^2|\mathbf{x}^{\mathbf{b}}\right)$ into Eq. (21), we obtain the BBF marginal likelihood under H_0 as

$$m_0^B(\mathbf{b}, \mathbf{x}) = \frac{k\left(\nu_\bullet\tau_\bullet^2\right)^{\frac{\nu_\bullet}{2}}\Gamma\left(\frac{n_1+n_2+\nu_\bullet-2}{2}\right)}{\pi^{\frac{n_1+n_2-2}{2}}\Gamma\left(\frac{\nu_\bullet}{2}\right)(n_1n_2)^{\frac{1}{2}}\left((n_1-1)s_1^2+(n_2-1)s_2^2+\nu_\bullet\tau_\bullet^2\right)^{\frac{n_1+n_2+\nu_\bullet-2}{2}}}, \tag{22}$$

with $\nu_\bullet$ and $\tau_\bullet^2$ as in Eq. (15), and k is an unspecified constant coming from the improper Jeffreys prior on the common mean parameter, $\pi^N(\boldsymbol{\mu})$ (similar to k_0 in Appendix A).

The marginal likelihoods under H_1 and H_2 are functions of the marginal likelihood under H_u, which is

$$m_u^B(\mathbf{b}, \mathbf{x}) = \frac{k\pi^{-\frac{n_1+n_2-2}{2}}(n_1n_2)^{-\frac{1}{2}}\left(\nu_\bullet\tau_\bullet^2\right)^{\nu_\bullet}\Gamma\left(\frac{n_1+\nu_\bullet-1}{2}\right)\Gamma\left(\frac{n_2+\nu_\bullet-1}{2}\right)}{\Gamma\left(\frac{\nu_\bullet}{2}\right)^2\left((n_1-1)s_1^2+\nu_\bullet\tau_\bullet^2\right)^{\frac{n_1+\nu_\bullet-1}{2}}\left((n_2-1)s_2^2+\nu_\bullet\tau_\bullet^2\right)^{\frac{n_2+\nu_\bullet-1}{2}}}, \tag{23}$$

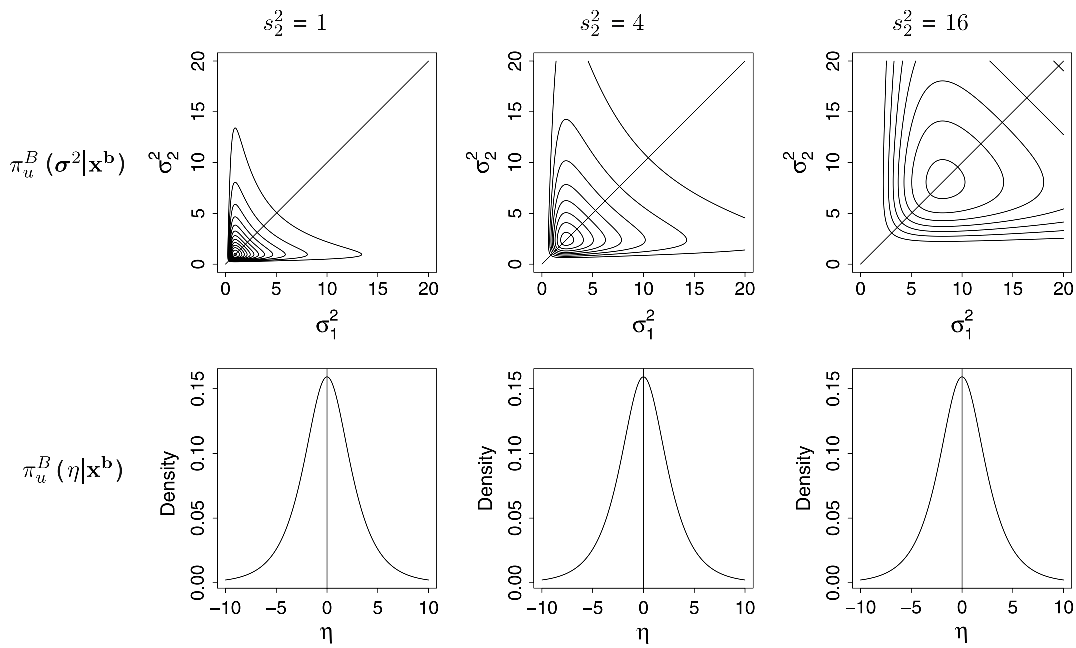


Fig. 3. The marginal unconstrained BBF prior $\pi_u^B\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right)$ (top row) and the corresponding prior $\pi_u^B\left(\eta = \log\left(\sigma_1^2/\sigma_2^2\right)|\mathbf{x}^\mathbf{b}\right)$ (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.075$. The prior $\pi_u^B\left(\eta|\mathbf{x}^\mathbf{b}\right)$ is always balanced.

with k as in Eq. (22). The marginal likelihoods under H_1 and H_2 are then given by

$$m_p^B(\mathbf{b}, \mathbf{x}) = \frac{P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)}{P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^\mathbf{b}\right)} m_u^B(\mathbf{b}, \mathbf{x})$$

$$= 2 \cdot P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right) \cdot m_u^B(\mathbf{b}, \mathbf{x}), \quad p = 1, 2, \tag{24}$$

with $P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^\mathbf{b}\right)$ as in Eq. (19), and the exact expression for $P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)$ is given in Eq. (B.3) in Appendix B. The derivation of Eqs. (22)–(24) follows steps similar to those in Appendix A. Note that the unspecified constant k cancels out in the computation of Bayes factors.

4.2.2. Evaluation of the method

We will now evaluate the BBF according to the five properties discussed in Section 3:

1. *Proper priors*. Eqs. (18) and (20), in combination with Eqs. (14)–(17), show that the priors on $\boldsymbol{\sigma}^2$ under H_0, H_1, and H_2 are proper (truncated) scaled-inverse-χ^2 distributions if $b_1n_1 + b_2n_2 > 2$.
2. *Minimal information*. As was set out in the previous section, the unconstrained prior is based on the assumption that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. A minimal study therefore consists of three observations, with at least one observation from each population. We can thus make the priors contain the information of a minimal study by setting $\mathbf{b} = (1.5/n_1, 1.5/n_2)'$. Note that this results in degrees of freedom of $\nu_\bullet = 1$ (see Eq. (15)).
3. *Scale invariance*. The BBF is scale-invariant for the same reason that the FBF is (see Section 4.1.2).
4. *Balancedness*. As was mentioned before, the unconstrained prior $\pi_u^B\left(\eta|\mathbf{x}^\mathbf{b}\right)$ is balanced by definition. An illustration is given in Fig. 3, which shows the priors on $\boldsymbol{\sigma}^2$ (top row) and η (bottom row) for sample variances $s_1^2 = 1$ and $s_2^2 \in \{1, 4, 16\}$, sample sizes $n_1 = n_2 = 20 = n$, and fractions $b_1 = b_2 = 1.5/n = 1.5/20 = 0.075$. It can be seen that $\pi_u^B\left(\eta|\mathbf{x}^\mathbf{b}\right)$ is always balanced.
5. *Occam's razor*. Fig. 4 shows the Bayes factors B_{1u}^B (solid line) and B_{2u}^B (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.075$. It can be seen that B_{1u}^B (B_{2u}^B) increases (decreases) monotonically as s_2^2 increases, favoring the more parsimonious inequality constrained hypothesis over the unconstrained hypothesis if the former is supported by the data. The Bayes factors thus function as Occam's razor. In fact, the Bayes factors go to 2 for very large and very small s_2^2, respectively, because H_1 and H_2 are twice as parsimonious as H_u.

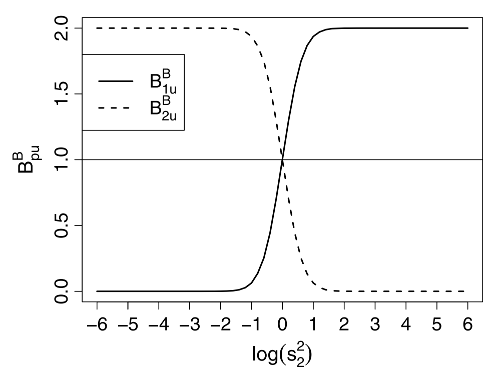


Fig. 4. Bayes factors B_{1u}^B (solid line) and B_{2u}^B (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.075$. The Bayes factors favor the more parsimonious inequality constrained hypothesis if it is supported by the data. This shows that B_{1u}^B and B_{2u}^B function as Occam's razor.

4.3. Adjusted fractional Bayes factor

Mulder (2014b) proposed a modification of the integration region in the FBF marginal likelihood under (in)equality constrained hypotheses to ensure that the latter always incorporates the complexity of an inequality constrained hypothesis. Compared to the FBF, the proposed modification is always larger for an inequality constrained hypothesis that is supported by the data. Even though this is essentially a good property, a possible disadvantage of this approach is that it results in a slight decrease of the evidence in favor of a true null hypothesis. For this reason we propose an alternative method in this article: we adjust the FBF marginal likelihood

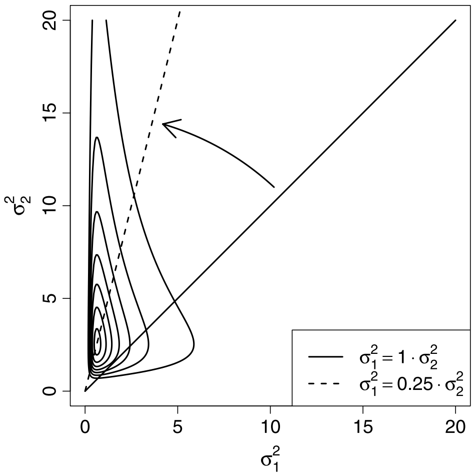


Fig. 5. Marginal unconstrained FBF prior $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right)$ for sample variances $s_1^2 = 1$ and $s_2^2 = 4$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. The probability mass above the line $\sigma_1^2 = a\sigma_2^2$, $a = 1$, is larger than that below it. We adjust the line by decreasing a until the probability mass above and below the line $\sigma_1^2 = a\sigma_2^2$ is equal to 0.5. For the depicted prior this is the case for $a = 0.25$.

under an inequality constrained hypothesis as suggested by Mulder (2014b), but we keep the marginal likelihood under the equality constrained hypothesis as in the FBF approach. We shall refer to this approach as the adjusted fractional Bayes factor (aFBF). We use the superscript aF to refer to the aFBF.

4.3.1. Marginal likelihoods

Following Mulder (2014b), we define the adjusted FBF marginal likelihood under an inequality constrained hypothesis as

$$m_p^{aF}(\mathbf{b}, \mathbf{x}) = \frac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)\pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu} d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2} f_u\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right)^\mathbf{b}\pi_u^N\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right) d\boldsymbol{\mu} d\boldsymbol{\sigma}^2},$$
$$p = 1, 2, \tag{25}$$

where $\mathbf{b} = (b_1, b_2)' \in (1/n_1, 1] \times (1/n_2, 1]$ as with the FBF. Note the two adjustments that were made compared to the standard FBF marginal likelihood given in Eq. (5). First, we use the unconstrained likelihood and Jeffreys prior. Second, in the denominator we integrate over an adjusted parameter space Ω_p^a, which will be defined shortly. We do not adjust the FBF marginal likelihoods under H_0 and H_u, that is, we set

$$m_0^{aF}(\mathbf{b}, \mathbf{x}) = m_0^F(\mathbf{b}, \mathbf{x}) \quad \text{and} \quad m_u^{aF}(\mathbf{b}, \mathbf{x}) = m_u^F(\mathbf{b}, \mathbf{x}). \tag{26}$$

The aFBF of H_p, $p = 1, 2$, against H_u is then given by

$$B_{pu}^{aF} = \frac{m_p^{aF}(\mathbf{b}, \mathbf{x})}{m_u^{aF}(\mathbf{b}, \mathbf{x})} = \frac{\int_{\Omega_p}\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}\right) d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right) d\boldsymbol{\sigma}^2}$$
$$= \frac{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)}{P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x}^\mathbf{b}\right)}, \tag{27}$$

where $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}\right)$ and $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right)$ are as in Eqs. (B.1) and (12), respectively. A derivation is given in Appendix D.

Now, we want $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x}^\mathbf{b}\right) = \int_{\Omega_p^a}\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right) d\boldsymbol{\sigma}^2 = 0.5$ (similar to $P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x}^\mathbf{b}\right)$ in Eq. (19)) to ensure that the automatic Bayes factor B_{pu}^{aF} functions as Occam's razor when evaluating an inequality constrained hypothesis. To achieve this, we define the adjusted parameter space Ω_p^a, $p = 1, 2$, as

$$\Omega_1^a := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < a\sigma_2^2\right\} \quad \text{and}$$
$$\Omega_2^a := \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > a\sigma_2^2\right\}, \tag{28}$$

where a is a constant chosen such that $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^\mathbf{b}\right) = P^F\left(\boldsymbol{\sigma}^2 \in \Omega_2^a|\mathbf{x}^\mathbf{b}\right) = 0.5$. Fig. 5 illustrates this. The plot shows $\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right)$ for sample variances $s_1^2 = 1$ and $s_2^2 = 4$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. Two lines $\sigma_1^2 = a\sigma_2^2$ are depicted, one for $a = 1$ and one for $a = 0.25$. To determine Ω_1^a and Ω_2^a we proceed as follows. It can be seen that the probability mass in Ω_1 (i.e. above the line $\sigma_1^2 = 1 \cdot \sigma_2^2$) is larger than that in Ω_2. By tuning a we tilt the line $\sigma_1^2 = a\sigma_2^2$ such that the probability mass above and below the line is equal to 0.5. For the prior depicted in Fig. 5 this is the case for $a = 0.25$. We thus have $\Omega_1^a = \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < 0.25 \cdot \sigma_2^2\right\}$ and $\Omega_2^a = \left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 > 0.25 \cdot \sigma_2^2\right\}$, and $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^\mathbf{b}\right) = P^F\left(\boldsymbol{\sigma}^2 \in \Omega_2^a|\mathbf{x}^\mathbf{b}\right) = 0.5$.

If we use $\mathbf{b} = (2/n_1, 2/n_2)'$ in order to satisfy the minimal information property, then it can be shown that $a = \frac{n_2(n_1-1)s_1^2}{n_1(n_2-1)s_2^2}$. In this case we can show that $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x}^\mathbf{b}\right) = 0.5$ by transforming the integral

$$P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^\mathbf{b}\right) = \int_{\Omega_1^a}\pi_u^F\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right) d\boldsymbol{\sigma}^2$$
$$= \int_{\left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < a\sigma_2^2\right\}} \text{Inv-}\chi^2\left(\sigma_1^2|\nu_1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|\nu_2, \tau_2^2\right) d\boldsymbol{\sigma}^2$$
$$= \int_{\left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\right\}} \text{Inv-}\chi^2\left(\sigma_1^2|1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|1, a\tau_2^2\right) d\boldsymbol{\sigma}^2$$
$$= \int_{\left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\right\}} \text{Inv-}\chi^2\left(\sigma_1^2|1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|1, \tau_1^2\right) d\boldsymbol{\sigma}^2$$
$$= \int_{\left\{\boldsymbol{\sigma}^2 \in \Omega_u : \sigma_1^2 < \sigma_2^2\right\}} \pi_u^{aF}\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right) d\boldsymbol{\sigma}^2 = 0.5, \tag{29}$$

with ν_j and τ_j^2, $j = 1, 2$, as in Eq. (13). Here we used the result that if $\sigma^2 \sim \text{Inv-}\chi^2\left(\nu, \tau^2\right)$, then $a\sigma^2 \sim \text{Inv-}\chi^2\left(\nu, a\tau^2\right)$. The density

$$\pi_u^{aF}\left(\boldsymbol{\sigma}^2|\mathbf{x}^\mathbf{b}\right) = \text{Inv-}\chi^2\left(\sigma_1^2|1, \tau_1^2\right) \text{Inv-}\chi^2\left(\sigma_2^2|1, \tau_1^2\right) \tag{30}$$

can be regarded as the implicit unconstrained prior in the aFBF approach. Note that irrespective of the exact choice of $\mathbf{b}$ there always exists an a that yields $P^F\left(\boldsymbol{\sigma}^2 \in \Omega_1^a|\mathbf{x}^\mathbf{b}\right) = P^F\left(\boldsymbol{\sigma}^2 \in \Omega_2^a|\mathbf{x}^\mathbf{b}\right) = 0.5$.

4.3.2. Evaluation of the method

We will now evaluate the aFBF according to the five properties discussed in Section 3:

1. *Proper priors*. As with the FBF, we must have $b_1 n_1 + b_2 n_2 > 2$ under H_0 and $b_j n_j > 1$, $j = 1, 2$, under H_1, H_2, and H_u to ensure that the priors are proper.
2. *Minimal information*. As was mentioned before, the minimal information property can be satisfied by setting $\mathbf{b} = \left(2/n_1, 2/n_2\right)'$.
3. *Scale invariance*. The aFBF is scale-invariant for the same reason that the FBF is (see Section 4.1.2).
4. *Balancedness*. In Eq. (30) we have seen that the implicit unconstrained prior on $\boldsymbol{\sigma}^2$ is a product of two scaled inverse-χ^2 distributions with identical hyperparameters. Thus the corresponding prior on η is balanced (see Appendix C).
5. *Occam's razor*. Fig. 6 shows the behavior of B_{1u}^{aF} (dotted line) as compared to B_{1u}^F (solid line) and B_{1u}^B (dashed line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$, sample sizes $n_1 = n_2 = 20$, and fractions $b_1 = b_2 = 0.1$. For $s_1^2 < s_2^2$ the Bayes factor B_{1u}^{aF} favors the more parsimonious inequality constrained hypothesis H_1: $\sigma_1^2 < \sigma_2^2$. It thus functions as Occam's razor.

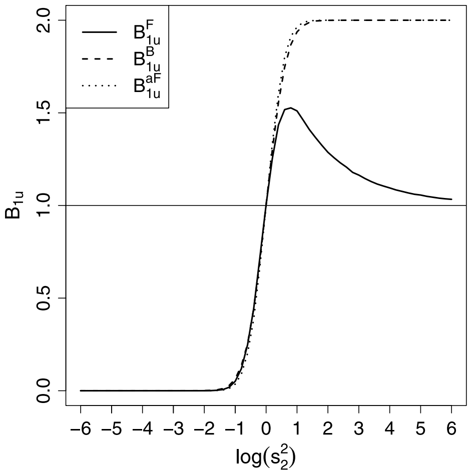


Fig. 6. Bayes factors B^F_{1u} (solid line), B^B_{1u} (dashed line), and B^{aF}_{1u} (dotted line) for sample variances $s_1^2 = 1$ and $s_2^2 \in [\exp(-6), \exp(6)]$ and sample sizes $n_1 = n_2 = 20$. In the FBF and the aFBF the fractions are $b_1 = b_2 = 0.1$, while in the BBF we have $b_1 = b_2 = 0.075$. For $s_1^2 < s_2^2$ the Bayes factor B^{aF}_{1u} favors the more parsimonious inequality constrained hypothesis $H_1 : \sigma_1^2 < \sigma_2^2$. It thus functions as Occam's razor.

5. Performance of the Bayes factors

We present results of a simulation study investigating the performance of the three automatic Bayes factors. We consider two normal populations $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$, where $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. That is, we consider four effect sizes $\sigma_2^2/\sigma_1^2 \in \{1.0, 1.5, 2.0, 2.5\}$. A study by Ruscio and Roche (2012, Table 2) indicates that these population variance ratios roughly correspond to {no, small, medium, large} effects in psychological research. We first investigate the strength of the evidence in favor of the true hypothesis H_t, $t = 0, 1$. The goal here is to see which automatic Bayes factor converges fastest to the true hypothesis. Following this, we consider frequentist error probabilities of selecting the wrong hypothesis. Note that from a Bayesian point of view these probabilities are of limited importance because Bayes factors are consistent in the sense that the evidence in favor of the true hypothesis grows to infinity as the sample size accumulates. These frequentist probabilities can be useful, however, to decide which automatic Bayes factor to use based on differences in error probability behavior.

5.1. Strength of evidence in favor of the true hypothesis

In this section we will investigate which automatic Bayes factor provides strongest evidence in favor of the true hypothesis. We shall use two measures of evidence. The first is the weight of evidence in favor of H_t against $H_{t'}$, where $t' = 1$ if $t = 0$ and $t' = 0$ otherwise. The weight of evidence is given by the logarithm of the Bayes factor, that is, $\log(B_{tt'})$. The second measure of evidence we use is the posterior probability of the true hypothesis. Assuming that all hypotheses are equally likely a priori (i.e. $P(H_0) = P(H_1) = P(H_2) = 1/3$, which is a standard default choice), it is given by $P(H_t|\mathbf{x}) = \frac{m_t(\mathbf{b},\mathbf{x})}{m_0(\mathbf{b},\mathbf{x})+m_1(\mathbf{b},\mathbf{x})+m_2(\mathbf{b},\mathbf{x})}$, where $m_t(\mathbf{b}, \mathbf{x})$ denotes the marginal likelihood under H_t. Both measures of evidence are computed for the FBF, the BBF, and the aFBF.

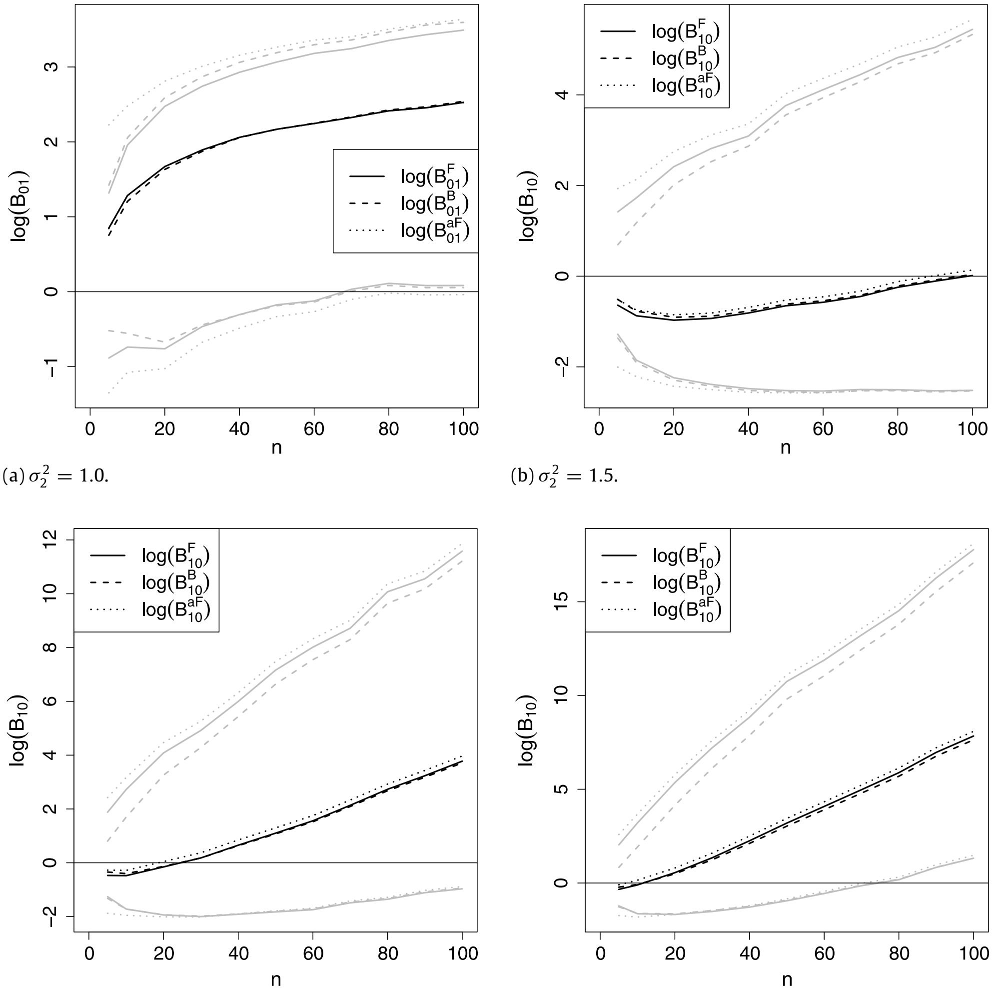
We drew 5000 samples of size $n_1 = n_2 = n \in \{5, 10, 20, \ldots, 100\}$ from X_1 and X_2. Denote these samples by $\mathbf{x}^{(m)} = \left(\mathbf{x}_1^{(m)}, \mathbf{x}_2^{(m)}\right)'$, $m = 1, \ldots, 5000$. For each $\mathbf{x}^{(m)}$ we computed the two measures of evidence $\log(B_{tt'})^{(m)}$ and $P\left(H_t|\mathbf{x}^{(m)}\right)$. Eventually, we computed the median of $\left\{\log(B_{tt'})^{(m)}\right\}_{m=1}^{5000}$ and $\left\{P\left(H_t|\mathbf{x}^{(m)}\right)\right\}_{m=1}^{5000}$ to estimate the average evidence in favor of H_t, as well as the 2.5%- and 97.5%-quantile to obtain an indication of the variability of the evidence.

Fig. 7 shows the results for the weight of evidence, $\log(B_{tt'})$. The plots show the median (black lines) and the 2.5%- and 97.5%-quantile (gray lines) as a function of the common sample size n for each $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. It can be seen that the three automatic Bayes factors provide similarly strong median evidence in favor of the true hypothesis (panels (a) to (d)). In panel (a) the dotted line for the aFBF is actually covered by the lines for the FBF and the BBF. If there is a positive effect (panels (b) to (d)), then the aFBF provides slightly stronger evidence in favor of the true hypothesis H_1 than the FBF and the BBF (as can be seen from the lines for the median and the 97.5%-quantile). The BBF, on the other hand, provides somewhat weaker evidence in favor of H_1. This is because the balanced prior slightly shrinks the posterior towards $\sigma_1^2 = \sigma_2^2$, which results in a loss of evidence in favor of an inequality constrained hypothesis that is supported by the data. The FBF and the aFBF are not affected by such shrinkage. Fig. 8 shows the simulation results for the posterior probability of the true hypothesis, $P(H_t|\mathbf{x})$. In the legends the superscripts F, B, and aF denote on which Bayes factor the posterior probability is based. The results are in line with those from Fig. 7. In fact, the advantage of the aFBF over the FBF and the BBF in terms of strength of evidence is a bit more pronounced. Overall, it can be concluded that the aFBF performs best: under H_0 it performs about as good as the FBF and the BBF, while under H_1 it slightly outperforms the latter two.

5.2. Frequentist error probabilities

Table 1 shows simulated frequentist error probabilities of the three automatic Bayes factors and the likelihood-ratio (LR) test for $\sigma_1^2 = 1$ and $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. For each σ_2^2 we drew 5000 samples of size $n_1 = n_2 = n \in \{5, 50, 500\}$ from $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$. On each sample we computed the Bayes factors and the LR test. In the Bayesian testing approach an error occurs if the true hypothesis H_t does not have the largest posterior probability, that is, if $P\left(H_{t'}|\mathbf{x}^{(m)}\right) > P\left(H_t|\mathbf{x}^{(m)}\right)$ for some $t' \neq t$. Here again we assumed equal prior probabilities of the hypotheses. In the frequentist approach an error occurs under H_0 if $p < \alpha$ and under H_1 if $p > \alpha \vee \left(p < \alpha \wedge s_1^2 > s_2^2\right)$. In the present simulation we set $\alpha = 0.05$. Table 1 shows the proportions of errors in the 5000 samples. It can be seen that the error probabilities of the three automatic Bayes factors are quite similar. Under H_0 the aFBF shows somewhat larger error probabilities. Under H_1, however, it has lower error probabilities than the FBF and the BBF, particularly for $n = 5$. Moreover, it can be seen that under H_1 the Bayes factors have lower error probabilities than the LR test. While the differences are considerable for $n = 5$, the LR test closes the gap as the sample size increases. One final remark concerns the error probabilities under H_0: While the LR test has unconditional error probabilities equal to $\alpha = 0.05$ regardless of the sample size, the conditional error probabilities of the three Bayes factors decrease as the sample size increases. This illustrates that the automatic Bayes factors are consistent whereas the p-value is not.

Additional insight into the performance of the three automatic Bayes factors is given in Table 2. It is well-known that p-values tend to overstate the evidence against the null hypothesis and that methods based on comparing likelihoods (such as Bayes factors and posterior probabilities of hypotheses) commonly yield weaker evidence against the null (see, for example, Berger & Sellke, 1987; Held, 2010; Sellke, Bayarri, & Berger, 2001). Table 2 shows that this also holds for the three automatic Bayes factors discussed in this article. The table can be read as follows. For sample sizes of $n_1 = n_2 = n = 5$ and sample variances of $s_1^2 = 1$ and $s_2^2 = 9.60$, the standard likelihood-ratio test of equality of variances yields a two-sided p-value of 0.05. The posterior probabilities of H_0 based on these sample data are $P^F(H_0|\mathbf{x}) = 0.26$, $P^B(H_0|\mathbf{x}) = 0.34$,



(a) $\sigma_2^2 = 1.0$.

(b) $\sigma_2^2 = 1.5$.

(c) $\sigma_2^2 = 2.0$.

(d) $\sigma_2^2 = 2.5$.

Fig. 7. Results of a simulation study investigating the performance of the FBF, the BBF, and the aFBF in testing variances of two normal populations $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$, where $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. The black lines depict the median weight of evidence in favor of the true hypothesis H_t, $\log(B_{tt'})$, as a function of the common sample size $n_1 = n_2 = n$. The gray lines depict the 2.5%- and 97.5%-quantile. It can be seen that if there is a positive effect (i.e. if $\sigma_1^2 < \sigma_2^2$), then the aFBF provides strongest evidence in favor of the true hypothesis H_1.

Table 1

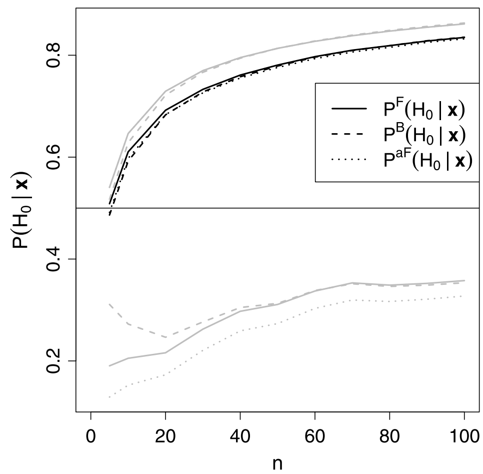
Frequentist error probabilities of the three automatic Bayes factors and the likelihood-ratio (LR) test for $\sigma_1^2 = 1, \sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$, and $n_1 = n_2 = n \in \{5, 50, 500\}$. In the LR test we set $\alpha = 0.05$. It can be seen that under H_1 the aFBF has lower error probabilities than the FBF and the BBF.

σ_2^2	1.0			1.5			2.0			2.5		
n	5	50	500	5	50	500	5	50	500	5	50	500
FBF	0.23	0.07	0.02	0.80	0.66	0.01	0.72	0.28	0.00	0.65	0.09	0.00
BBF	0.26	0.07	0.02	0.79	0.66	0.01	0.69	0.28	0.00	0.62	0.09	0.00
aFBF	0.36	0.08	0.02	0.72	0.63	0.01	0.60	0.26	0.00	0.54	0.08	0.00
LR test	0.05	0.05	0.05	0.94	0.71	0.00	0.92	0.33	0.00	0.89	0.11	0.00

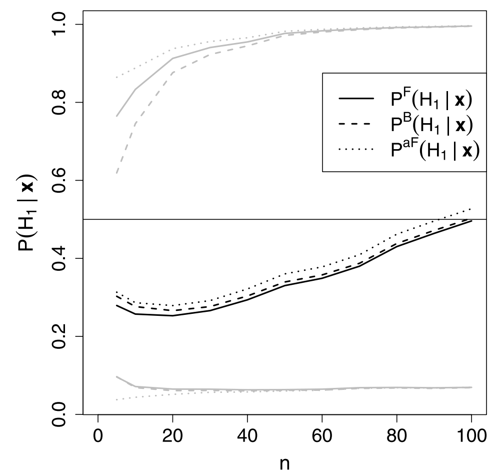
Table 2

Comparison of two-sided p-values and posterior probabilities of H_0, denoted by $P(H_0|\mathbf{x})$. The superscripts F, B, and aF denote on which Bayes factor $P(H_0|\mathbf{x})$ is based. For example, sample sizes of $n_1 = n_2 = n = 5$ and sample variances of $s_1^2 = 1.00$ and $s_2^2 = 9.60$ yield a p-value of 0.05 and posterior probabilities of H_0 of 0.26, 0.34, and 0.19. It can be seen that while the p-values indicate evidence against H_0, the posterior probabilities tell us that H_0 is quite likely given the sample data.

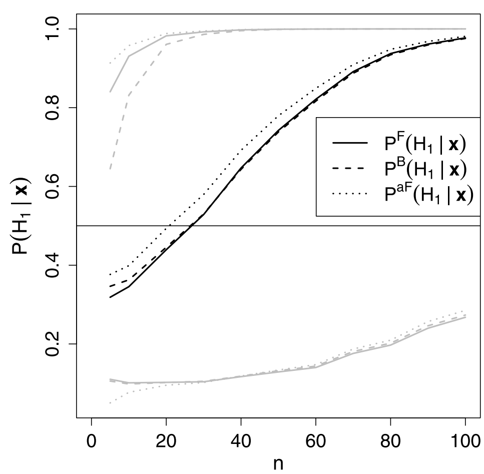
n	s_1^2	$p = 0.05$				$p = 0.01$			
		s_2^2	$P^F(H_0\|\mathbf{x})$	$P^B(H_0\|\mathbf{x})$	$P^{aF}(H_0\|\mathbf{x})$	s_2^2	$P^F(H_0\|\mathbf{x})$	$P^B(H_0\|\mathbf{x})$	$P^{aF}(H_0\|\mathbf{x})$
5	1.00	9.60	0.26	0.34	0.19	23.15	0.11	0.28	0.07
10	1.00	4.03	0.29	0.34	0.23	6.54	0.11	0.20	0.08
20	1.00	2.53	0.34	0.36	0.29	3.43	0.13	0.16	0.10
50	1.00	1.76	0.43	0.43	0.39	2.11	0.17	0.18	0.14
100	1.00	1.49	0.51	0.50	0.48	1.69	0.21	0.21	0.19



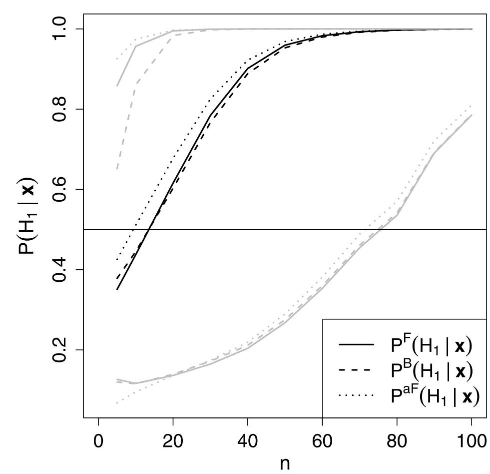
(a) $\sigma_2^2 = 1.0$.



(b) $\sigma_2^2 = 1.5$.



(c) $\sigma_2^2 = 2.0$.



(d) $\sigma_2^2 = 2.5$.

Fig. 8. Results of a simulation study investigating the performance of the FBF, the BBF, and the aFBF in testing variances of two normal populations $X_1 \sim \mathrm{N}(0, 1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2^2)$, where $\sigma_2^2 \in \{1.0, 1.5, 2.0, 2.5\}$. The black lines depict the median posterior probability of the true hypothesis H_t, $P(H_t|\mathbf{x})$, as a function of the common sample size $n_1 = n_2 = n$. The gray lines depict the 2.5%- and 97.5%-quantile. In the legends the superscripts F, B, and aF denote on which Bayes factor the posterior probability is based. It can be seen that if there is a positive effect (i.e. if $\sigma_1^2 < \sigma_2^2$), then the aFBF provides strongest evidence in favor of the true hypothesis H_1.

and $P^{aF}(H_0|\mathbf{x}) = 0.19$. From the frequentist significance test we would thus conclude that there is evidence against H_0, whereas the posterior probabilities tell us that there is some evidence for H_0 given the observed data. This discrepancy between the p-value and the posterior probabilities of H_0 becomes even more pronounced for larger sample sizes. A similar picture emerges for $p = 0.01$: While the p-value tells us that there is strong evidence against H_0, it is difficult to rule out H_0 given posterior probabilities roughly between 0.1 and 0.3. It can be seen that the posterior probabilities of H_0 decrease as the p-value decreases. This suggests that only very small p-values should be considered indicative of evidence against H_0, particularly if sample sizes are large.

6. Empirical data examples

In this section we apply the three automatic Bayes factors to two empirical data sets.

6.1. Example 1: variability of intelligence in children (Arden & Plomin, 2006)

We first consider a study by Arden and Plomin (2006) investigating differences in variance of intelligence between girls and boys. Psychological research has consistently found males to be more variable in intellectual abilities than females (e.g. Feingold, 1992). Arden and Plomin therefore assumed that this finding would also apply to children. Their dependent variable of interest was a general ability factor extracted from several tests of verbal and non-verbal ability. The authors expected that boys would show larger variance on this factor than girls, which can be formulated in the hypothesis $H_1: \sigma_f^2 < \sigma_m^2$, where σ_f^2 and σ_m^2 denote the population variances of females and males, respectively. The competing hypotheses are $H_0: \sigma_f^2 = \sigma_m^2$ and $H_2: \sigma_f^2 > \sigma_m^2$.

In samples of $n_f = 1366$ girls and $n_m = 1136$ boys of age 10, Arden and Plomin found sample variances of $s_f^2 = 0.92$ and $s_m^2 = 1.10$. Table 3 provides the Bayes factors B_{10} and B_{12} and the posterior probabilities of H_0, H_1, and H_2 (assuming equal prior probabilities) for these sample data. As can be seen, the posterior probabilities of H_0, H_1, and H_2 are approximately 0.13, 0.87, and 0.00 for all three automatic Bayes factors. An immediate conclusion we can draw from these results is that we can basically rule out H_2. The Bayes factors B_{10} and B_{12}, and the posterior probability of H_1, $P(H_1|\mathbf{x})$, indicate positive evidence in favor of H_1. However, the evidence does not appear to be strong enough to completely rule out H_0. The two-sided p-value for these data obtained from the standard likelihood-ratio test equals 0.002, which would commonly be interpreted as sufficient evidence to reject H_0 in favor of the two-sided alternative.

Table 3
Results for two empirical data examples.

	Example 1					Example 2				
	B_{10}	B_{12}	$P(H_0\|\mathbf{x})$	$P(H_1\|\mathbf{x})$	$P(H_2\|\mathbf{x})$	B_{01}	B_{02}	$P(H_0\|\mathbf{x})$	$P(H_1\|\mathbf{x})$	$P(H_2\|\mathbf{x})$
FBF	6.32	1176.58	0.14	0.86	0.00	7.14	5.52	0.76	0.10	0.14
BBF	6.43	1261.63	0.13	0.87	0.00	7.73	4.96	0.75	0.10	0.15
aFBF	6.68	1316.52	0.13	0.87	0.00	7.21	5.47	0.76	0.10	0.14

6.2. Example 2: precision of burn wound assessments (Martin, Lundy, & Rickard, 2014)

We next reanalyze data from a study by Martin et al. (2014) investigating the precision of burn wound assessments by UK Armed Forces medical personnel. The percentage of the total body surface area that is burned (%TBSA burned) is a very important measure in the treatment of burn victims. The authors had two groups of medical personnel estimate the %TBSA burned for one particular burn case. The first group consisted of $n_1 = 20$ experienced burn specialists, while the second group consisted of $n_2 = 40$ relatively inexperienced participants of a surgical training course. Martin et al. expected the experienced burn specialists to be less variable in their %TBSA burned estimates than the inexperienced medical personnel. This expectation can be formulated in the hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2$, the competing hypotheses being $H_0\colon \sigma_1^2 = \sigma_2^2$ and $H_2\colon \sigma_1^2 > \sigma_2^2$.

Martin et al. found sample variances of $s_1^2 = 105.88$ and $s_2^2 = 100.60$. The two-sided p-value obtained from the standard likelihood-ratio test equals $p = 0.86$ for these sample data. From this p-value it can be concluded that there is not enough evidence to reject the null hypothesis that the two groups are equally heterogeneous. However, we cannot conclude that there is evidence in favor of the null hypothesis since p-values do not imply this kind of information. The p-value of 0.86 thus leaves us in a state of ignorance. The Bayes factor on the other hand can be used to quantify the relative evidence in favor of a null hypothesis. Table 3 provides the Bayes factors B_{01} and B_{02} and the posterior probabilities of H_0, H_1, and H_2 (assuming equal prior probabilities). The Bayes factors and the posterior probability of H_0, $P(H_0|\mathbf{x})$, indicate positive evidence in favor of H_0. In particular, the posterior probability of H_0 is approximately 0.76 for all three automatic Bayes factors. However, the posterior probabilities of H_1 and H_2 are between 0.10 and 0.15, indicating that it is difficult to completely rule out either of the two hypotheses based on the sample data.

7. Discussion

In this article we presented three automatic Bayes factors for testing variances of two independent normal distributions: the FBF, the BBF, and the aFBF. The three Bayes factors are fully automatic and thus readily applicable. All the user needs to provide is the two sample sizes and sample variances. This makes the Bayes factors particularly valuable for both statisticians and applied researchers who are interested in a user-friendly Bayesian method for testing two variances.

The methods were theoretically evaluated on the basis of five properties: proper priors, minimal information, scale invariance, balancedness, and Occam's razor. As was shown, the FBF satisfies neither the balancedness property nor the Occam's razor property when testing inequality constraints on variances. The BBF and the aFBF, on the other hand, satisfy all five properties. In the BBF, an automatic balanced prior is constructed based on equal prior distributions for the variances with minimal information. In the aFBF the FBF marginal likelihood is adjusted such that it adequately incorporates the parsimony of an inequality constrained hypothesis. The simulation study indicates that the aFBF provides strongest evidence in favor of a true inequality constrained hypothesis. The slightly worse performance of the BBF is caused by the fact that the balanced prior shrinks the posterior towards the boundary of the constrained space where $\sigma_1^2 = \sigma_2^2$, resulting in weaker evidence in favor of a true inequality constrained hypothesis. The FBF and the aFBF, on the other hand, are not affected by prior shrinkage.

One possible point of debate relating to all three Bayes factors is the choice of the fraction b. In this article we used the minimal information approach to specifying b, which is a widely accepted principle. However, there are other approaches to specifying b, each of which pursuing a different goal (see, for example, Conigliani & O'Hagan, 2000; O'Hagan, 1995). All important formulas in this article are expressed in terms of b. It is therefore straightforward to use a different b if desired.

There are two natural extensions of our approach to testing two variances. First, we are often interested in testing variances of $J > 2$ independent populations. Relevant hypotheses in this case include the null hypothesis $H_0\colon \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_J^2$, the order constrained hypothesis $H_1\colon \sigma_1^2 < \sigma_2^2 < \cdots < \sigma_J^2$, and hypotheses with combinations of equality and inequality constraints, for example $H_2\colon \sigma_1^2 = \sigma_2^2 < \cdots < \sigma_J^2$. The second natural extension is testing variances of dependent populations, which is relevant when analyzing repeated measurement data. Based on our findings the automatic Bayes factors discussed in this paper will prove useful for these more complex testing problems.

Acknowledgments

We would like to thank Marcel van Assen, the editor Eric-Jan Wagenmakers, and two anonymous reviewers for their helpful comments which improved the manuscript. This research was partly supported by a Rubicon grant which was awarded to the second author by the Netherlands Organisation for Scientific Research (NWO).

Appendix A. Derivation of $m_0^F(\mathbf{b}, \mathbf{x})$

Plugging $f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)$, $f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)^{\mathbf{b}}$, and $\pi_0^N\left(\boldsymbol{\mu},\sigma^2\right)$ into Eq. (5) gives us

$$m_0^F(\mathbf{b},\mathbf{x}) = \frac{\int_{\Omega_0}\int_{\mathbb{R}^2} f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)\pi_0^N\left(\boldsymbol{\mu},\sigma^2\right)d\boldsymbol{\mu}d\sigma^2}{\int_{\Omega_0}\int_{\mathbb{R}^2} f_0\left(\mathbf{x}|\boldsymbol{\mu},\sigma^2\right)^{\mathbf{b}}\pi_0^N\left(\boldsymbol{\mu},\sigma^2\right)d\boldsymbol{\mu}d\sigma^2} = \frac{m_0^F(\mathbf{x})}{m_0^F\left(\mathbf{x}^{\mathbf{b}}\right)}. \tag{A.1}$$

We first derive the denominator $m_0^F\left(\mathbf{x}^{\mathbf{b}}\right)$. Note that the Jeffreys prior can be written as $\pi_0^N\left(\boldsymbol{\mu},\sigma^2\right) = k_0\sigma^{-2}$, where k_0 is an unspecified normalizing constant (see, for instance, O'Hagan, 1995). Then

$$\begin{aligned} m_0^F\left(\mathbf{x}^{\mathbf{b}}\right) &= \int_{\mathbb{R}^+}\int_{\mathbb{R}^2} f\left(\mathbf{x}_1|\mu_1,\sigma^2\right)^{b_1} f\left(\mathbf{x}_2|\mu_2,\sigma^2\right)^{b_2} k_0\sigma^{-2}d\boldsymbol{\mu}d\sigma^2 \\ &= k_0\int_{\mathbb{R}^+}\int_{\mathbb{R}} f\left(\mathbf{x}_1|\mu_1,\sigma^2\right)^{b_1}d\mu_1 \\ &\quad\times\int_{\mathbb{R}} f\left(\mathbf{x}_2|\mu_2,\sigma^2\right)^{b_2}d\mu_2\,\sigma^{-2}d\sigma^2 \end{aligned}$$

$$= k_0 \int_{\mathbb{R}^+} (b_1 n_1)^{-\frac{1}{2}} \left(\sigma^2 2\pi\right)^{-\frac{b_1 n_1 - 1}{2}} \times \exp\left(-\frac{b_1 (n_1-1) s_1^2}{2\sigma^2}\right) (b_2 n_2)^{-\frac{1}{2}} \left(\sigma^2 2\pi\right)^{-\frac{b_2 n_2 - 1}{2}} \times \exp\left(-\frac{b_2 (n_2-1) s_2^2}{2\sigma^2}\right) \sigma^{-2} d\sigma^2$$
$$= k_0 (b_1 b_2)^{-\frac{1}{2}} (n_1 n_2)^{-\frac{1}{2}} (2\pi)^{-\frac{b_1 n_1 + b_2 n_2 - 2}{2}} \times \int_{\mathbb{R}^+} \left(\sigma^2\right)^{-\left(\frac{b_1 n_1 + b_2 n_2 - 2}{2} + 1\right)} \times \exp\left(-\frac{b_1 (n_1-1) s_1^2 + b_2 (n_2-1) s_2^2}{2\sigma^2}\right) d\sigma^2$$
$$= k_0 (b_1 b_2)^{-\frac{1}{2}} (n_1 n_2)^{-\frac{1}{2}} \times \pi^{-\frac{b_1 n_1 + b_2 n_2 - 2}{2}} \Gamma\left(\frac{b_1 n_1 + b_2 n_2 - 2}{2}\right) \times \left(b_1 (n_1-1) s_1^2 + b_2 (n_2-1) s_2^2\right)^{-\frac{b_1 n_1 + b_2 n_2 - 2}{2}}, \quad \text{(A.2)}$$

where the integrand in the last but one line is the kernel of a scaled inverse-χ^2 distribution with parameters $\nu = b_1 n_1 + b_2 n_2 - 2$ and $\tau^2 = \frac{b_1(n_1-1)s_1^2 + b_2(n_2-1)s_2^2}{b_1 n_1 + b_2 n_2 - 2}$. We obtain $m_0^F(\mathbf{x})$ by setting $b_1 = b_2 = 1$ in the expression for $m_0^F(\mathbf{x^b})$. Dividing $m_0^F(\mathbf{x})$ by $m_0^F(\mathbf{x^b})$ eventually yields the expression given in Eq. (8). Note that the unspecified constant k_0 cancels out in this step.

Appendix B. Probability that σ^2 is in Ω_p

For the FBF we have

$$P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p | \mathbf{x}\right) = \int_{\Omega_p} \text{Inv-}\chi^2\left(\sigma_1^2 | n_1 - 1, s_1^2\right) \times \text{Inv-}\chi^2\left(\sigma_2^2 | n_2 - 1, s_2^2\right) d\boldsymbol{\sigma}^2, \quad p = 1, 2, \quad \text{(B.1)}$$

and

$$P^F\left(\boldsymbol{\sigma}^2 \in \Omega_p | \mathbf{x^b}\right) = \int_{\Omega_p} \text{Inv-}\chi^2\left(\sigma_1^2 | b_1 n_1 - 1, \frac{b_1 (n_1-1) s_1^2}{b_1 n_1 - 1}\right) \times \text{Inv-}\chi^2\left(\sigma_2^2 | b_2 n_2 - 1, \frac{b_2 (n_2-1) s_2^2}{b_2 n_2 - 1}\right) d\boldsymbol{\sigma}^2, \quad p = 1, 2. \quad \text{(B.2)}$$

For the BBF we have

$$P^B\left(\boldsymbol{\sigma}^2 \in \Omega_p | \mathbf{x}\right) = \int_{\Omega_p} \text{Inv-}\chi^2\left(\sigma_1^2 | n_1 + \nu_\bullet - 1, \frac{(n_1-1) s_1^2 + \nu_\bullet \tau_\bullet^2}{n_1 + \nu_\bullet - 1}\right) \times \text{Inv-}\chi^2\left(\sigma_2^2 | n_2 + \nu_\bullet - 1, \frac{(n_2-1) s_2^2 + \nu_\bullet \tau_\bullet^2}{n_2 + \nu_\bullet - 1}\right) d\boldsymbol{\sigma}^2, \quad p = 1, 2. \quad \text{(B.3)}$$

The integrals cannot be solved analytically, but they can be approximated numerically using Monte Carlo methods: first we draw samples from the two scaled inverse-χ^2 distributions Inv-$\chi^2\left(\sigma_1^2|\nu_1, \tau_1^2\right)$ and Inv-$\chi^2\left(\sigma_2^2|\nu_2, \tau_2^2\right)$. An approximation of the integral is then given by the proportion of draws that fall in Ω_p.

Appendix C. Distribution of $\eta = \log(\sigma_1^2/\sigma_2^2)$

Let $\pi_{\sigma_1^2,\sigma_2^2}\left(\sigma_1^2, \sigma_2^2\right) = \text{Inv-}\chi^2_{\sigma_1^2}\left(\sigma_1^2|\nu_1, \tau_1^2\right) \text{Inv-}\chi^2_{\sigma_2^2}\left(\sigma_2^2|\nu_2, \tau_2^2\right)$ be the joint distribution of σ_1^2 and σ_2^2. We derive the distribution of $\eta = \log\left(\sigma_1^2/\sigma_2^2\right)$. To do so we first derive the distribution of $\zeta = \sigma_1^2/\sigma_2^2$ and subsequently apply the transformation $\eta = \log(\zeta)$. To determine the distribution of ζ we first determine the joint distribution of ζ and σ_2^2, which is given by

$$\pi_{\zeta,\sigma_2^2}\left(\zeta, \sigma_2^2\right) = \left|\frac{\partial \sigma_1^2}{\partial \zeta}\right| \pi_{\sigma_1^2,\sigma_2^2}\left(\sigma_1^2, \sigma_2^2\right) = \frac{\left(\frac{\nu_1 \tau_1^2}{2}\right)^{\frac{\nu_1}{2}} \left(\frac{\nu_2 \tau_2^2}{2}\right)^{\frac{\nu_2}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \zeta^{-\left(\frac{\nu_1}{2}+1\right)} \left(\sigma_2^2\right)^{-\left(\frac{\nu_1+\nu_2}{2}+1\right)} \times \exp\left(-\frac{\nu_1 \tau_1^2/\zeta + \nu_2 \tau_2^2}{2\sigma_2^2}\right). \quad \text{(C.1)}$$

We then obtain the marginal distribution of ζ by integrating out σ_2^2:

$$\pi_\zeta(\zeta) = \int_{\mathbb{R}^+} \pi_{\zeta,\sigma_2^2}\left(\zeta, \sigma_2^2\right) d\sigma_2^2 = \frac{\left(\nu_1 \tau_1^2\right)^{\frac{\nu_1}{2}} \left(\nu_2 \tau_2^2\right)^{\frac{\nu_2}{2}} \Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \times \zeta^{-\left(\frac{\nu_1}{2}+1\right)} \left(\nu_1 \tau_1^2 \zeta^{-1} + \nu_2 \tau_2^2\right)^{-\frac{\nu_1+\nu_2}{2}}. \quad \text{(C.2)}$$

Eventually, we obtain the distribution of η by applying the transformation $\eta = \log(\zeta)$:

$$\pi_\eta(\eta) = \left|\frac{d \exp(\eta)}{d\eta}\right| \pi_\zeta(\exp(\eta)) = \frac{\left(\nu_1 \tau_1^2\right)^{\frac{\nu_1}{2}} \left(\nu_2 \tau_2^2\right)^{\frac{\nu_2}{2}} \Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \times \exp(-\eta)^{\frac{\nu_1}{2}} \left(\nu_1 \tau_1^2 \exp(-\eta) + \nu_2 \tau_2^2\right)^{-\frac{\nu_1+\nu_2}{2}}. \quad \text{(C.3)}$$

Next, we show that $\pi_\eta(\eta)$ is balanced if and only if $\pi_{\sigma_1^2,\sigma_2^2}\left(\sigma_1^2, \sigma_2^2\right) = \text{Inv-}\chi^2_{\sigma_1^2}\left(\sigma_1^2|\nu, \tau^2\right) \text{Inv-}\chi^2_{\sigma_2^2}\left(\sigma_2^2|\nu, \tau^2\right)$. In other words, we show that $\pi_\eta(\eta)$ is symmetric about 0 and nonincreasing in $|\eta|$ if and only if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$. We first use a proof by contrapositive to show that if $\pi_\eta(\eta)$ is balanced, then $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$. Assume that $\neg\left(\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2\right)$. We show that $\pi_\eta(\eta)$ is not balanced. We consider three cases.

Case 1. $\nu_1 \neq \nu_2 \wedge \tau_1^2 \neq \tau_2^2$. Note that if $\pi_\eta(\eta)$ was symmetric about 0 and nonincreasing in $|\eta|$, then it would have a single mode at 0. In this case we would have $\frac{d}{d\eta}\pi_\eta(0) = 0$. However, if $\nu_1 \neq \nu_2 \wedge \tau_1^2 \neq \tau_2^2$, then $\frac{d}{d\eta}\pi_\eta(0) \neq 0$, which shows that $\pi_\eta(\eta)$ is not balanced.

Case 2. $\nu_1 = \nu_2 \wedge \tau_1^2 \neq \tau_2^2$. Note that if $\pi_\eta(\eta)$ was symmetric about 0, then we would have $\pi_\eta(\eta) = \pi_\eta(-\eta)$. Let $\nu_1 = \nu_2 = \nu$. Then

$$\pi_\eta(\eta) = \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2} \left(\tau_1^2 \tau_2^2\right)^{\frac{\nu}{2}} \left(\tau_1^2 \exp\left(-\frac{\eta}{2}\right) + \tau_2^2 \exp\left(\frac{\eta}{2}\right)\right)^{-\nu} \neq \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2} \left(\tau_1^2 \tau_2^2\right)^{\frac{\nu}{2}} \left(\tau_1^2 \exp\left(\frac{\eta}{2}\right) + \tau_2^2 \exp\left(-\frac{\eta}{2}\right)\right)^{-\nu} = \pi_\eta(-\eta), \quad \text{(C.4)}$$

since $\tau_1^2 \neq \tau_2^2$. This shows that $\pi_\eta(\eta)$ is not symmetric about 0, and thus not balanced.

Case 3. $\nu_1 \neq \nu_2 \wedge \tau_1^2 = \tau_2^2$. The argument is analogous to that in Case 2.

We next show that if $\nu_1 = \nu_2 \wedge \tau_1^2 = \tau_2^2$, then $\pi_\eta(\eta)$ is balanced. Let $\nu_1 = \nu_2 = \nu$ and $\tau_1^2 = \tau_2^2 = \tau^2$. Then $\pi_\eta(\eta)$ is symmetric about

0 since

$$\pi_\eta(\eta) = \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2}\left(\exp\left(-\frac{\eta}{2}\right)+\exp\left(\frac{\eta}{2}\right)\right)^{-\nu} = \frac{\Gamma(\nu)}{\Gamma\left(\frac{\nu}{2}\right)^2}\left(\exp\left(\frac{\eta}{2}\right)+\exp\left(-\frac{\eta}{2}\right)\right)^{-\nu} = \pi_\eta(-\eta). \quad \text{(C.5)}$$

Eventually, note that $\exp(-\eta/2)+\exp(\eta/2)$ is strictly monotonically increasing for $|\eta| = \eta > 0$, in which case the inverse of this expression and $\pi_\eta(\eta)$ are strictly monotonically decreasing (i.e. nonincreasing).

Appendix D. Derivation of B_{pu}^{aF}

$$B_{pu}^{aF} = \frac{m_p^{aF}(\mathbf{b},\mathbf{x})}{m_u^{aF}(\mathbf{b},\mathbf{x})} = \frac{m_p^{aF}(\mathbf{b},\mathbf{x})}{m_u^{F}(\mathbf{b},\mathbf{x})}$$
$$= \frac{\dfrac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)^{\mathbf{b}}\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}}{\dfrac{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)^{\mathbf{b}}\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}}$$
$$= \frac{\dfrac{\int_{\Omega_p}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}}{\dfrac{\int_{\Omega_p^a}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)^{\mathbf{b}}\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)^{\mathbf{b}}\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2}}$$
$$= \frac{\int_{\Omega_p}\int_{\mathbb{R}^2} \dfrac{f_u(\mathbf{x}|\mu,\sigma^2)\pi_u^N(\mu,\sigma^2)}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2} d\boldsymbol{\mu} d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2} \dfrac{f_u(\mathbf{x}|\mu,\sigma^2)^{\mathbf{b}}\pi_u^N(\mu,\sigma^2)}{\int_{\Omega_u}\int_{\mathbb{R}^2} f_u(\mathbf{x}|\mu,\sigma^2)^{\mathbf{b}}\pi_u^N(\mu,\sigma^2)d\mu d\sigma^2} d\boldsymbol{\mu} d\boldsymbol{\sigma}^2}$$
$$= \frac{\int_{\Omega_p}\int_{\mathbb{R}^2} \pi_u^F(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x})\, d\boldsymbol{\mu} d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a}\int_{\mathbb{R}^2} \pi_u^F(\boldsymbol{\mu},\boldsymbol{\sigma}^2|\mathbf{x^b})\, d\boldsymbol{\mu} d\boldsymbol{\sigma}^2}$$
$$= \frac{\int_{\Omega_p} \pi_u^F(\boldsymbol{\sigma}^2|\mathbf{x})\, d\boldsymbol{\sigma}^2}{\int_{\Omega_p^a} \pi_u^F(\boldsymbol{\sigma}^2|\mathbf{x^b})\, d\boldsymbol{\sigma}^2} = \frac{P^F(\boldsymbol{\sigma}^2 \in \Omega_p|\mathbf{x})}{P^F(\boldsymbol{\sigma}^2 \in \Omega_p^a|\mathbf{x^b})}. \quad \text{(D.1)}$$

References

Arden, R., & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences*, *41*(1), 39–48.

Berger, J. O., & Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, *2*(3), 317–335.

Berger, J. O., & Mortera, J. (1999). Default Bayes factors for nonnested hypothesis testing. *Journal of the American Statistical Association*, *94*(446), 542–554.

Berger, J. O., & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, *91*(433), 109–122.

Berger, J. O., & Pericchi, L. R. (2001). Objective Bayesian methods for model selection: Introduction and comparison. In P. Lahiri (Ed.), *Model selection* (pp. 135–207). Beachwood, OH: Institute of Mathematical Statistics.

Berger, J. O., & Sellke, T. (1987). Testing a point null hypothesis: The irreconcilability of *P* values and evidence. *Journal of the American Statistical Association*, *82*(397), 112–122.

Borkenau, P., Hřebíčková, M., Kuppens, P., Realo, A., & Allik, J. (2013). Sex differences in variability in personality: A study in four samples. *Journal of Personality*, *81*(1), 49–60.

Conigliani, C., & O'Hagan, A. (2000). Sensitivity of the fractional Bayes factor to prior distributions. *Canadian Journal of Statistics*, *28*(2), 343–352.

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research*, *62*(1), 61–84.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Held, L. (2010). A nomogram for *P* values. *BMC Medical Research Methodology*, *10*(1), 21.

Hubbard, R., & Armstrong, J. S. (2006). Why we don't really know what statistical significance means: Implications for educators. *Journal of Marketing Education*, *28*(2), 114–120.

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

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, *10*(4), 477–493.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*(1–2), 187–192.

Martin, N. A. J., Lundy, J. B., & Rickard, R. F. (2014). Lack of precision of burn surface area calculation by UK Armed Forces medical personnel. *Burns*, *40*(2), 246–250.

Mulder, J. (2014a). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology*, *67*(1), 153–171.

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

Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, *46*(2), 1–39.

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*(4), 887–906.

Mulder, J., & Wagenmakers, E.-J. in preparation, Editors' introduction to the special issue 'Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments'. *Journal of Mathematical Psychology*.

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *57*(1), 99–138.

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 & Review*, *16*(2), 225–237.

Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology*, *8*(1), 1–11.

Sellke, T., Bayarri, M. J., & Berger, J. O. (2001). Calibration of *p* values for testing precise null hypotheses. *The American Statistician*, *55*(1), 62–71.

Spiegelhalter, D. J., & Smith, A. F. M. (1982). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *44*(3), 377–387.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*(6), 1057–1064.