Bayesian Evaluation of Constrained Hypotheses on Variances of Multiple Independent Groups

Böing-Messing, F.; van Assen, M.A.L.M.; Hofman, A.D.; Hoijtink, H.; Mulder, J.

DOI
10.1037/met0000116

Publication date
2017

Document Version
Final published version

Published in
Psychological Methods

Psychological Methods
2017, Vol. 22, No. 2, 262–287

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000116

Bayesian Evaluation of Constrained Hypotheses on Variances of Multiple Independent Groups

Florian Böing-Messing
Tilburg University

Marcel A. L. M. van Assen
Tilburg University and Utrecht University

Abe D. Hofman
University of Amsterdam

Herbert Hoijtink
Utrecht University and CITO Institute for Educational Measurement, Arnhem, the Netherlands

Joris Mulder
Tilburg University

Abstract

Research has shown that independent groups often differ not only in their means, but also in their variances. Comparing and testing variances is therefore of crucial importance to understand the effect of a grouping variable on an outcome variable. Researchers may have specific expectations concerning the relations between the variances of multiple groups. Such expectations can be translated into hypotheses with inequality and/or equality constraints on the group variances. Currently, however, no methods are available for testing (in)equality constrained hypotheses on variances. This article proposes a novel Bayesian approach to this challenging testing problem. Our approach has the following useful properties: First, it can be used to simultaneously test multiple (non)nested hypotheses with equality as well as inequality constraints on the variances. Second, our approach is fully automatic in the sense that no subjective prior specification is needed. Only the hypotheses need to be provided. Third, a user-friendly software application is included that can be used to perform this Bayesian test in an easy manner.

Translational Abstract

Data analysis in the psychological sciences commonly focuses on averages. However, by disregarding the variability of the observations one runs the risk of overlooking crucial information in the data. In fact, there are often reasons to expect a certain structure of the variability across groups of different people. For example, one would expect observations from a treatment group to be more variable than observations from a control group because subjects react differently to the treatment. Such an expectation can be translated into the hypothesis "the treatment group is more variable than the control group." To test this hypothesis one needs to compare it to a competing hypothesis. A possible competitor is the hypothesis "the treatment and the control group are equally variable." In this article we use Bayesian statistics to test such hypotheses about the structure of the variability across two or more groups. The results of a simulation study indicate that our method is able to detect the correct hypothesis if the sample size is large enough. We present a user-friendly software application that can be used to perform our Bayesian test in a relatively easy manner. An application of our testing procedure to data from the Math Garden online learning environment (https://www.mathsgarden.com/) shows that our method provides valuable answers to research questions concerning the structure of the variability across groups.

Keywords: Bayes factor, heterogeneity, heteroscedasticity, homogeneity of variance, inequality constraint

Supplemental materials: http://dx.doi.org/10.1037/met0000116.supp

Florian Böing-Messing, Department of Methodology and Statistics, Tilburg University; Marcel A. L. M. van Assen, Department of Methodology and Statistics, Tilburg University, and Department of Sociology, Utrecht University; Abe D. Hofman, Department of Psychology, Psychological Methods, University of Amsterdam; Herbert Hoijtink, Department of Methodology and Statistics, Utrecht University, and CITO Institute for Educational Measurement, Arnhem, the Netherlands; Joris Mulder, Department of Methodology and Statistics, Tilburg University.

We thank Axel Mayer from RWTH Aachen University for helping us create the Shiny application. This research was partly supported by a Rubicon grant which was awarded to Joris Mulder by the Netherlands Organisation for Scientific Research (NWO). The statistical methods discussed in this article were presented at the 2016 World Meeting of the International Society for Bayesian Analysis (https://bayesian.org/) and at the 2015 Winter Conference of the Interuniversity Graduate School of Psychometrics and Sociometrics (http://www.iops.nl/). The data from the illustrative example are part of a larger data set containing data from the Math Garden. Multiple articles have been published presenting parts of this larger data set. The particular data set used in the present article as well as the analyses and results from the illustrative example were not presented prior to publication of this article.

Correspondence concerning this article should be addressed to Florian Böing-Messing, Department of Methodology and Statistics, Tilburg University, Postbus 90153, 5000 LE Tilburg, the Netherlands. E-mail: florian.boeingmessing@gmail.com

Data analysis in psychological research commonly focuses on measures of central tendency such as means and regression coefficients. Measures of dispersion like variances receive relatively little attention. By disregarding the dispersion, however, researchers run the risk of overlooking vital information in the data. Carroll (2003) distinguished two situations in which it is crucial to carefully consider the structure of variances. The first is the situation in which the variability systematically depends on known factors. An example is heteroscedasticity in ANOVA and regression. For instance, it has been pointed out that in experimental studies treatments often not only affect group means, but also group variances (e.g., Bryk & Raudenbush, 1988; Grissom, 2000; Ruscio & Roche, 2012). However, heterogeneity of variances is common in existing groups as well (e.g., Grissom, 2000; Ruscio & Roche, 2012). For example, males have been found to be more variable than females on a variety of measures (e.g., Lehre, Lehre, Laake, & Danbolt, 2009). Furthermore, it is frequently observed that the variability changes systematically with time (e.g., Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Hultsch, MacDonald, & Dixon, 2002). For example, a method that allows for the variability to systematically depend on known factors is the beta regression approach of Smithson and Verkuilen (2006). The authors model the mean as well as the variance as a function of (possibly different) predictors, thus treating the variance as a parameter of interest rather than as a nuisance parameter. The second situation in which variances play a crucial role is in multilevel modeling. Here researchers need to carefully model the variability at multiple levels, which results in multiple variance components. For example, Verhagen and Fox (2013) proposed a test on variance components in multilevel IRT models to check for measurement invariance in cross-national surveys. Furthermore, Kim and Seltzer (2011) examined heterogeneity in residual variance in multilevel models applied to (quasi-)experimental data in order to detect differential response to treatments. In the present article the focus is on heterogeneity of variances in one-way ANOVA designs with independent groups in the first situation.

There are often reasons to expect a certain structure of the variances of multiple independent groups. Typically one expects that certain groups are more heterogeneous than others, less heterogeneous, or equally heterogeneous. Such expectations can be translated into equality and inequality constrained hypotheses on the group variances. For example, in experimental studies one would expect treatment groups to have larger variances than control groups because participants respond differently to treatments (e.g., Bryk & Raudenbush, 1988; Grissom, 2000). Suppose we compare a control group with two treatment groups receiving a mild and an intense treatment, respectively. A conceivable hypothesis in this case would be $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$, where σ_1^2 is the variance of the control group and σ_2^2 and σ_3^2 are the variances of the groups receiving the mild and the intense treatment, respectively. Note that H_1 states that the intense treatment produces larger variance than the mild treatment. To see whether there is evidence in favor of H_1 we test it against one or more competing hypotheses. Potential competitors are the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ stating equality of variances and the complement of H_1 given by H_2 : not $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$. The complement H_2 entails all possible hypotheses except H_1. Hence, testing an order constrained hypothesis like H_1 against its complement tells us whether there is evidence in favor of our expected order or whether another hypothesis is more likely. Note that the interest is solely on the group variances, whereas the group means are treated as nuisance parameters.

Theoretical considerations often suggest (in)equality constrained hypotheses on the variances of existing groups as well. For example, Aunola et al. (2004) hypothesized that the variance of mathematical abilities either increases or decreases across grades. For $J \geq 2$ grades this can be expressed in the two competing hypotheses $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2 : \sigma_J^2 < \cdots < \sigma_1^2$, where σ_j^2 denotes the variance in grade j. The idea behind an increase (H_1) is that students who start out with high mathematical potential develop their mathematical abilities faster than students with low potential, which increases interindividual differences. A decrease in the variability of mathematical abilities (H_2) might occur because systematic instruction at school helps students with low mathematical potential catch up, so that interindividual differences decrease. Another potential competing hypothesis would be the null hypothesis $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$. Note that H_1 and H_2 are in agreement with models of development over time. For example, in the random slope model variances may increase over time, decrease over time, or first decrease and then increase over time (Snijders & Bosker, 2012). Constrained hypotheses on the variances of existing groups are conceivable in a variety of psychological research areas. For example, research on gender differences often finds males to be more variable in their intellectual abilities and personality than females (e.g., Borkenau, Hřebíčková, Kuppens, Realo, & Allik, 2013; Feingold, 1992). Gerontological studies have found that the variability of reaction times increases with age (e.g., Hultsch et al., 2002). Research on psychological disorders has shown that ADHD patients tend to be more variable in their attentional performances than groups of people who do not suffer from ADHD (e.g., Silverstein, Como, Palumbo, West, & Osborn, 1995). Furthermore, research on person-in-context behavior suggests that the variability of people's behavior may differ across situations. For example, Van Mechelen (2009) argued that in an aggression context the variability may depend on the amount of social control in a situation, where high social control results in homogeneous behavior and thus low variability.

The standard approach to testing variances is null hypothesis significance testing (NHST). Classical NHST procedures like the likelihood ratio test or Levene's test (Levene, 1960) test the null hypothesis stating that all J variances are equal, $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$, against the alternative hypothesis stating that the variances are not all equal, H_a : not $\sigma_1^2 = \cdots = \sigma_J^2$. In testing the order constrained hypothesis $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ using classical NHST procedures one would proceed as follows: First we test the null against the alternative hypothesis. If we are able to reject the null hypothesis, we check whether the sample variances follow the order stated in the order constrained hypothesis. For more than two groups this is done by pairwise comparisons. This approach entails two problems: First, it suffers from Type I error inflation if we do not adjust the significance level for multiple testing. If we do adjust the significance level, then the procedure suffers from low power (e.g., Cohen, 1992). Second, it is possible that the pairwise comparisons produce contradictory results (e.g., $H_0 : \sigma_1^2 = \sigma_2^2$ and $H_0 : \sigma_1^2 = \sigma_3^2$ are not rejected, but $H_0 : \sigma_2^2 = \sigma_3^2$ is).

Motivated by these disadvantages, Gastwirth, Gel, and Miao (2009) proposed an NHST procedure for testing the null hypothesis against an order constrained hypothesis. The advantage of this test is that it has higher power to detect an order effect. However, the method does not allow testing the null against an alternative hypothesis with a combination of equality and inequality constraints on the variances. This is a serious limitation given the large number of distinct hypotheses we can formulate. Using different combinations of equality and inequality constraints, we can specify dozens of distinct hypotheses on three variances. For more than three groups there are well over 100 distinct hypotheses. Furthermore, the test by Gastwirth et al. (2009) does not solve the problems inherent in all NHST procedures: First, NHST procedures are not able to quantify evidence in favor of a hypothesis, no matter whether it is a null, an order constrained, or an unconstrained hypothesis (e.g., Wagenmakers, 2007). Second, it often happens that researchers have multiple competing hypotheses they would like to compare. NHST procedures do not allow testing these hypotheses against one another to determine which is most supported by the data. All one can do is test each hypothesis against the null, which does not answer the research question which hypothesis receives strongest support.

Given the problems with NHST procedures, it seems natural to use an information criterion like the Akaike information criterion (AIC; Akaike, 1973) or the Bayesian information criterion (BIC; Schwarz, 1978) to compare the hypotheses. However, these criteria cannot be used to test inequality constrained hypotheses. Both the AIC and the BIC involve a penalty term that measures the complexity of a hypothesis by the number of parameters. However, under inequality constrained hypotheses the number of parameters is not a suitable measure of the complexity because each inequality constraint effectively reduces the complexity. For example, the order constrained hypothesis $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ is less complex than the unconstrained hypothesis $H_u : \sigma_1^2, \sigma_2^2, \sigma_3^2$ because under H_1 the variances can take on fewer values (e.g., σ_1^2 cannot be greater than σ_2^2). As a solution to this problem, Anraku (1999) proposed the order-restricted information criterion (ORIC). However, the ORIC is designed for testing order constrained hypotheses on means. At this point it is unclear whether this methodology can be generalized to the case of testing equality and inequality constrained hypotheses on variances. Note that the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002) and the Watanabe-Akaike information criterion (WAIC; Watanabe, 2010) do not provide a solution to this problem because they do not properly take the parsimony introduced by inequality constraints into account (Mulder et al., 2009; Gelman, Hwang, & Vehtari, 2014). Under certain conditions the DIC and the WAIC are asymptotically equal to leave-one-out cross-validation (Gelman et al., 2014), which implies that the latter is not suitable for testing inequality constrained hypotheses on variances either.

In this article we adopt a Bayesian approach to testing equality and inequality constrained hypotheses on variances using Bayes factors (Jeffreys, 1961; Kass & Raftery, 1995). The Bayes factor is a Bayesian hypothesis testing and model selection criterion. It provides a solution to the aforementioned problems inherent in NHST procedures and existing information criteria. In particular, the Bayes factor quantifies the evidence in favor of a hypothesis. This holds for all types of hypotheses: The Bayes factor allows quantification of evidence in favor of a null hypothesis, order constrained hypotheses, and hypotheses with a combination of equality and inequality constraints. Furthermore, using the Bayes factor it is straightforward to simultaneously test multiple hypotheses against one another. In this case the Bayes factor tells us which hypothesis is most supported by the data. Bayes factors have a number of additional desirable properties: First, contrary to NHST procedures, Bayes factors do not require the hypotheses under consideration to be nested (e.g., Berger & Mortera, 1999). Bayes factors are therefore able to directly test, for example, $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_2 : \sigma_3^2 < \sigma_2^2 < \sigma_1^2$. Second, Bayes factors automatically function as Occam's razor. This means that if two hypotheses describe the data equally well, then the Bayes factor automatically chooses the more parsimonious hypothesis. This is a useful property of Bayes factors because it is frequently observed that parsimonious hypotheses that describe the data well are more likely to be correct than complex ones. Third, Bayes factors are consistent. This means that the Bayes factor always chooses the true hypothesis if we have enough data.

Bayes factors have been developed for various testing problems frequently encountered in the psychological sciences. For instance, Rouder, Speckman, Sun, Morey, and Iverson (2009) proposed a Bayesian *t* test. Klugkist, Laudy, and Hoijtink (2005) discussed a Bayes factor for testing hypotheses on mean parameters in analysis of variance designs. Mulder, Hoijtink, and Klugkist (2010) presented methods for Bayesian testing of means and regression coefficients in the multivariate normal linear model. Gu, Mulder, Deković, and Hoijtink (2014) proposed an approximate Bayes factor for evaluating hypotheses with inequality constraints on means and regression parameters. In the present article we propose a novel Bayes factor for testing equality and inequality constrained hypotheses on variances of multiple independent groups. Our methodology builds upon the fractional Bayes factor of O'Hagan (1995) in combination with the prior adjustment of Mulder (2014b) and Böing-Messing and Mulder (2016).

The remainder of this article is structured as follows. First, we discuss the statistical model and options for formulating hypotheses on the group variances. We then give a brief introduction to the Math Garden (Klinkenberg, Straatemeier, & van der Maas, 2011; Straatemeier, 2014), which we use to illustrate the importance of testing (in)equality constrained hypotheses on variances. Next, we discuss Bayes factors for testing hypotheses on variances. We first apply the fractional Bayes factor (O'Hagan, 1995) to the testing problem and show that it may not function as Occam's razor when testing inequality constrained hypotheses. As a novel solution to this problem we propose an adjusted fractional Bayes factor. The performance of the new method is illustrated in a simulation study. Following this, we continue the illustrative example by applying the adjusted fractional Bayes factor to data from the Math Garden. We then present a user-friendly software application for computing the adjusted fractional Bayes factor. We conclude the article with a discussion of our approach.

Model and Hypotheses

We consider the one-way ANOVA design with $J \geq 2$ independent groups of size n_j, $j = 1, \ldots, J$. Each observation in group j is assumed to be independent and normally distributed with mean μ_j and variance σ_j^2. The unconstrained likelihood with no constraints on the group means and variances is given by

$$f_u(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{j=1}^{J} \prod_{i=1}^{n_j} N(x_{ij} \mid \mu_j, \sigma_j^2), \tag{1}$$

where $\mathbf{x}$ are the data, x_{ij} is the ith observation from the jth group, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_J)'$ is the vector of group means, and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_J^2)'$ is the vector of group variances.

Hypotheses on the variances can be formulated using two basic types of constraints: equality constraints and inequality constraints. With equality constraints we can specify equalities of two or more variances, for example $H : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$. Inequality constraints are used to formulate expectations regarding differences in magnitude between variances, for example $H : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$. If we do not expect certain relations between variances, then we simply do not impose constraints on them. We shall use the comma symbol (,) to indicate that there are no constraints between variances, for example $H_u : \sigma_1^2, \sigma_2^2, \sigma_3^2$. We refer to the hypothesis with no constraints on the variances as the unconstrained hypothesis. In formulating hypotheses we may combine equality constraints, inequality constraints, and no constraints between variances, for example $H : \sigma_1^2 = \sigma_2^2 < \sigma_3^2, \sigma_4^2$. Another hypothesis that is often of interest is the complement of an order constrained hypothesis. For example, the complement of the order constrained hypothesis $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ is given by H_2 : not $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$, for which we also write H_2 : not H_1 in short. The complement entails all possible hypotheses except the order constrained hypothesis. We may also test the complement of multiple orders. For example, Aunola et al. (2004) expected the variance of mathematical abilities to either increase or decrease across grades. This corresponds to the two order constrained hypotheses $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ and $H_2 : \sigma_J^2 < \cdots < \sigma_1^2$, for which the complement is given by H_3 : not (H_1 or H_2). Note that one may also perform the classical test of the null hypothesis $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$ against the unconstrained alternative hypothesis $H_u : \sigma_1^2, \ldots, \sigma_J^2$ if the interest is on whether the group variances are equal or not. The likelihood under a constrained hypothesis H_t is a truncation of the unconstrained likelihood in Equation (1) in the parameter space that is admissible under H_t, which we denote by Ω_t:

$$f_t(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = f_u(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) I_{\Omega_t}(\boldsymbol{\sigma}^2), \tag{2}$$

where $I_{\Omega_t}(\boldsymbol{\sigma}^2)$ is an indicator function that equals 1 if the variances $\boldsymbol{\sigma}^2$ are in the admissible parameter space Ω_t and 0 if the variances are outside the admissible parameter space.

Illustrative Example: The Math Garden

The Math Garden (Klinkenberg et al., 2011; Straatemeier, 2014) is an online adaptive learning environment for basic mathematics. It is currently used by more than 300,000 children in primary education, involving more than 4,000 schools. Next to providing children and teachers with an online learning tool, the system opens up a valuable database for researchers. In this article we present an analysis of children's abilities in four different games, each covering one of the basic mathematical operations addition, subtraction, multiplication, and division.

As mentioned in the introduction, Aunola et al. (2004) hypothesized that the variance of mathematical abilities either increases or decreases across grades. This suggests testing the following two research hypotheses in the Math Garden:

$$\begin{aligned} &H_1 : \sigma_1^2 < \cdots < \sigma_J^2, \\ &H_2 : \sigma_J^2 < \cdots < \sigma_1^2, \end{aligned} \tag{3}$$

where σ_j^2 is the variance of mathematical abilities in grade j and J is the number of grades to be compared. Thus, H_1 states an increase in variance, whereas H_2 states a decrease. We shall test these two research hypotheses against two competing hypotheses:

$$\begin{aligned} &H_0 : \sigma_1^2 = \cdots = \sigma_J^2, \\ &H_3 : \text{ not } (H_0 \text{ or } H_1 \text{ or } H_2). \end{aligned} \tag{4}$$

Here H_0 is the classical null hypothesis stating equality of variances. Hypothesis H_3 is the complement of H_0, H_1, and H_2. We include it to cover all possible hypotheses in case neither the research hypotheses nor the null hypothesis is supported by the data. In the Math Garden a player's ability is estimated separately for each of the four games addition, subtraction, multiplication, and division. That is, each player has a separate ability estimate for each game they play. We will therefore test the hypotheses in Equations (3) and (4) for each game separately.

Bayes Factors for Testing Constrained Hypotheses on Variances

The Bayes factor is a Bayesian testing criterion that can be used to quantify the relative evidence in the data between two hypotheses. The main ingredient of the Bayes factor is the marginal likelihood of the data under each hypothesis. The marginal likelihood of the data $\mathbf{x}$ under the constrained hypothesis H_t, denoted by m_t, is defined by the integral over the product of the likelihood, denoted by f_t, and the prior, denoted by π_t, over the admissible parameter space under H_t. The marginal likelihood can be expressed as

$$m_t(\mathbf{x}) = \int_{\Omega_t} \int_{\mathbb{R}^J} f_t(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \pi_t(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2, \tag{5}$$

where the likelihood f_t under the constrained hypothesis H_t was given in Equation (2), the prior distribution π_t contains the information about the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}^2$ before observing the data, which will be discussed below, and Ω_t denotes the constrained parameter space of the variances under H_t. For example, for $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ the constrained space Ω_1 corresponds to the subspace of the variances that is in agreement with the ordering $\sigma_1^2 < \cdots < \sigma_J^2$. The marginal likelihood quantifies how well the model and the prior under H_t were able to predict the observed data (Jeffreys, 1961; Kass & Raftery, 1995).

In order to determine the evidence in the data in favor of a hypothesis, say H_1, relative to another hypothesis, say H_2, the ratio of the marginal likelihoods needs to be computed via

$$B_{12} = \frac{m_1(\mathbf{x})}{m_2(\mathbf{x})}, \tag{6}$$

which is known as the Bayes factor of hypothesis H_1 against hypothesis H_2. If the Bayes factor B_{12} is larger (smaller) than 1, this indicates that the evidence in the data in favor of H_1 (H_2) is stronger than the evidence in favor of H_2 (H_1). For example, a Bayes factor of $B_{12} = 10$ implies that the evidence in the data in favor of H_1 is 10 times as strong as the evidence in favor of H_2. Kass and Raftery (1995) provided interpretation guidelines for the Bayes factor as stated in Table 1. We would like to emphasize,

Table 1
Interpretation Guidelines for the Bayes Factor B_{12} Testing Hypothesis H_1 Against Hypothesis H_2 (From Kass & Raftery, 1995)

B_{12}	Evidence in favor of H_1
1 to 3	Not worth more than a bare mention
3 to 20	Positive
20 to 150	Strong
>150	Very strong

however, that these guidelines should not be used as strict rules when interpreting Bayes factors. A researcher should decide for himself or herself whether a Bayes factor of, say, $B_{12} = 120$ is enough to completely rule out hypothesis H_2 in comparison with hypothesis H_1.

Prior specification is an important step when computing the marginal likelihood. First, it is important to note that priors should not be specified in an ad hoc manner because the Bayes factor strongly depends on the exact choice of the prior. For instance, the Bayes factor for a null hypothesis against an unconstrained alternative hypothesis can be made arbitrarily large when specifying the prior under the unconstrained alternative extremely vague. This is known as Bartlett's phenomenon (e.g., Bartlett, 1957; Jeffreys, 1961; Liang, Paulo, Molina, Clyde, & Berger, 2008; Lindley, 1957). Alternatively, one might consider using noninformative improper priors, which are commonly used in objective Bayesian estimation (Berger, 2006). When using Bayes factors, however, it is not possible to work with noninformative improper priors because these contain undefined normalizing constants which do not cancel out when computing the marginal likelihoods and Bayes factors according to Equations (5) and (6).

Thus, in order to quantify the relative evidence in the data between constrained hypotheses on variances using the Bayes factor one needs to carefully formulate proper priors for the unknown parameters under all hypotheses under consideration. For instance, in the Math Garden example a proper prior needs to be specified for the group variances under H_1 satisfying the increasing order, the group variances under H_2 satisfying the decreasing order, the common group variance under H_0, and the group variances under the complement hypothesis H_3. Because often precise prior information about the degree of heterogeneity across populations is not available, specification of proper priors is a difficult task for a researcher. This holds especially when testing hypotheses with constraints on variances.

To avoid this limitation statisticians have developed automatic (or default) marginal likelihoods and Bayes factors that enable researchers to automatically quantify the relative evidence in the data between the hypotheses. These default Bayes factors can be computed in an automatic fashion without needing to specify proper priors for the model parameters based on one's subjective prior beliefs. Well-known examples are the fractional Bayes factor (O'Hagan, 1995), the intrinsic Bayes factor (Berger & Pericchi, 1996), and the Bayes factor based on expected-posterior priors (Mulder et al., 2009; Pérez & Berger, 2002). Here we shall focus on the fractional Bayes factor because it is computationally efficient and has desirable theoretical properties (O'Hagan, 1995, 1997).

Fractional Bayes Factors

The fractional Bayes factor (FBF) was proposed by O'Hagan (1995) to circumvent the need to specify a proper prior based on external prior information. In the FBF, the marginal likelihood is defined as

$$m_t^F(\mathbf{x}, b) = \frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}, \tag{7}$$

where π_t^N denotes a noninformative improper prior and the fraction b can take on values between 0 and 1. Thus, the marginal likelihood in the FBF corresponds to the standard marginal likelihood based on a noninformative improper prior divided by the standard marginal likelihood where the likelihood is raised to the power of the fraction b. O'Hagan (1995) motivates this form of the marginal likelihood in the context of partial Bayes factors. In particular, he argues that the fraction of the likelihood (i.e., the likelihood to the power of b) contains a part of the information in the full likelihood in the sense that the fraction of the likelihood is approximately equal to the likelihood based on a training sample if we set $b = m/n$, where both the sample size n and the training sample size m are large. As will be elaborated below, the fraction b controls the amount of information in the implicit automatic proper prior.

The noninformative improper prior we use in Equation (7) is the standard independence Jeffreys prior. For a constrained hypothesis, this noninformative improper prior is proportional to the product of the reciprocals of the unique variances truncated in the inequality constrained parameter space (if there are inequality constraints present). For example, under $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$ with one unique variance, say, σ^2, and $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ with J unique variances that are inequality constrained, the noninformative improper priors are given by

$$\begin{aligned}&\pi_0^N(\boldsymbol{\mu}, \sigma^2) = C_0 \times \sigma^{-2} \quad \text{and}\\ &\pi_1^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) = C_1 \times \sigma_1^{-2} \times \cdots \times \sigma_J^{-2} \times I(\sigma_1^2 < \cdots < \sigma_J^2),\end{aligned} \tag{8}$$

respectively, where $I(\sigma_1^2 < \cdots < \sigma_J^2)$ is an indicator function that equals 1 if $\sigma_1^2 < \cdots < \sigma_J^2$ and 0 otherwise, and C_0 and C_1 denote the respective undefined normalizing constants. Because the noninformative improper prior appears in the numerator as well as in the denominator in the marginal likelihood in Equation (7), the undefined constants in the improper prior cancel out in the FBF approach. Note that the noninformative priors imply flat priors for the group means.

The fraction b controls how much of the information in the data is used to specify an automatic proper prior. This can be made explicit by rewriting the marginal likelihood in Equation (7) following Gilks (1995):

$$\begin{aligned}&m_t^F(\mathbf{x}, b)\\ &= \int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^{1-b} \frac{f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)}{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}d\boldsymbol{\sigma}^2} d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2\\ &= \int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^{1-b}\pi_t(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\,|\,\mathbf{x}^b)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2,\end{aligned} \tag{9}$$

where

$$\pi_t(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mid \mathbf{x}^b) = \frac{f_t(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b \pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)}{\int_{\Omega_t} \int_{\mathbb{R}^J} f_t(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b \pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} \tag{10}$$

is the automatic proper prior that is obtained by updating the noninformative improper prior with a fraction b of the likelihood. Note that the symbol $\mathbf{x}^b$ is used to illustrate that this prior contains a fraction b of the information in the complete data $\mathbf{x}$. As can be seen from Equation (9), in computing the marginal likelihood the fraction b of the likelihood is used to obtain a proper automatic prior and the remaining fraction 1 – b is used for hypothesis testing. It is generally recommendable to choose the fraction b based on the minimal number of observations that is needed to obtain a proper automatic prior when updating the improper Jeffreys prior (e.g., Berger & Mortera, 1999; O'Hagan, 1995). In our testing problem with 2J unknown parameters (i.e., J unknown means and J unknown variances), we need at least 2J observations to obtain a proper prior when updating the improper Jeffreys prior. This implies setting $b = 2J/N$, where $N = \sum_{j=1}^{J} n_j$ is the total sample size. This choice ensures that the remaining fraction 1 – b that is used for hypothesis testing is maximal. As was shown by O'Hagan (1995), the FBF is consistent under very general settings, which implies that as the sample size grows to infinity, the evidence in favor of the true hypothesis goes to infinity. If we use a minimal fraction the evidence in favor of the true hypothesis goes fastest to infinity, which makes the minimal fraction the optimal choice.

Fractional Bayes Factors for an Inequality Constrained Test

Next we apply the FBF to test the inequality constrained hypothesis $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against the unconstrained hypothesis $H_u : \sigma_1^2, \sigma_2^2, \sigma_3^2$. The inequality constrained subspace under H_1 can be written as $\Omega_1 = \{\boldsymbol{\sigma}^2 \mid \sigma_1^2 < \sigma_2^2 < \sigma_3^2\}$. As shown in Appendix A, the FBF for H_1 against H_u can be written as the posterior probability that the constraints of H_1 hold divided by the automatic prior probability that the constraints of H_1 hold:

$$B_{1u}^F = \frac{m_1^F(\mathbf{x}, b)}{m_u^F(\mathbf{x}, b)} = \frac{P(\boldsymbol{\sigma}^2 \in \Omega_1 \mid \mathbf{x})}{P(\boldsymbol{\sigma}^2 \in \Omega_1 \mid \mathbf{x}^b)} = \frac{P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 \mid \mathbf{x})}{P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 \mid \mathbf{x}^b)}. \tag{11}$$

The unconstrained marginal automatic prior for the variances, which is needed to compute the probability in the denominator in Equation (11), can be obtained by integrating the group means out of the joint automatic prior:

$$\pi_u(\boldsymbol{\sigma}^2 \mid \mathbf{x}^b) = \int_{\mathbb{R}^3} \pi_u(\boldsymbol{\mu}, \boldsymbol{\sigma}^2 \mid \mathbf{x}^b) d\boldsymbol{\mu} = \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2 \,\middle|\, bn_j - 1, \frac{b(n_j - 1)s_j^2}{bn_j - 1}\right), \tag{12}$$

which is a product of scaled inverse-χ^2 distributions with degrees of freedom of $bn_j - 1$ and scale hyperparameters of $\frac{b(n_j-1)s_j^2}{bn_j-1}$, where $s_j^2 = \frac{1}{n_j-1}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2$ is the sample variance of group j. In this setting the minimal fraction is given by $b = 6/N$, where $N = n_1 + n_2 + n_3$. The unconstrained marginal posterior can simply be obtained by plugging $b = 1$ into Equation (12), which yields

$$\pi_u(\boldsymbol{\sigma}^2 \mid \mathbf{x}) = \prod_{j=1}^{3} \text{Inv-}\chi^2(\sigma_j^2 \mid n_j - 1, s_j^2). \tag{13}$$

The distributions above can be used to obtain a large sample of, say, S = 100,000 draws from the unconstrained posterior and unconstrained automatic prior (see Gelman, Carlin, Stern, & Rubin, 2004, for information on how to sample from the scaled inverse-χ^2 distribution). Subsequently, by taking the proportion of unconstrained draws that satisfy the constraints of H_1, the fractional Bayes factor in Equation (11) can be computed as

$$B_{1u}^F \approx \frac{S^{-1}\sum_{s=1}^{S} I(\sigma_{1,post}^{2(s)} < \sigma_{2,post}^{2(s)} < \sigma_{3,post}^{2(s)})}{S^{-1}\sum_{s=1}^{S} I(\sigma_{1,prior}^{2(s)} < \sigma_{2,prior}^{2(s)} < \sigma_{3,prior}^{2(s)})}, \tag{14}$$

where $\boldsymbol{\sigma}_{post}^{2(s)} = (\sigma_{1,post}^{2(s)}, \sigma_{2,post}^{2(s)}, \sigma_{3,post}^{2(s)})'$ and $\boldsymbol{\sigma}_{prior}^{2(s)} = (\sigma_{1,prior}^{2(s)}, \sigma_{2,prior}^{2(s)}, \sigma_{3,prior}^{2(s)})'$ are the sth draw from the unconstrained posterior and automatic prior, respectively, for $s = 1, \ldots, S$.

It is important to note that the use of a common fraction b for all groups may be problematic in the case of unbalanced data with unequal group sizes. For example, when $n_1 = 10$, $n_2 = 20$, and $n_3 = 30$, it holds that $b = 6/60 = 0.1$. This results in prior degrees of freedom of 0, 1, and 2, for σ_1^2, σ_2^2, and σ_3^2, respectively. However, the degrees of freedom must be larger than 0. This shows that the standard FBF approach is not generally applicable in the case of unequal group sizes. We come back to this issue in the next section.

Another important consequence of using a fraction of the data for constructing the automatic prior in Equation (12) is that the scale hyperparameter of each variance σ_j^2 depends on the corresponding sample variance s_j^2. This implies that the automatic prior is concentrated around the observed effect, which has undesirable consequences when testing inequality constrained hypotheses on variances. We illustrate this with an example. For the moment, let us consider a balanced data set with equal group sizes of $n_j = n = 20$, for $j = 1, 2$, and 3, and let the sample variances satisfy $s_1^2 = 1$, $s_2^2 = s$, and $s_3^2 = s^2$. Thus, if $s > 1$, then there is evidence in favor of H_1 because the sample variances are in agreement with the inequality constraints under H_1. Similarly, if $s < 1$, then there is evidence against H_1 because the sample variances are not in agreement with the inequality constraints. Note that the degrees of freedom in the automatic prior equal $6/60 \times 20 - 1 = 1$, which implies a distribution with minimal information.

Figure 1 shows the FBF for H_1 against H_u (solid line) when letting s^2 increase from $\exp(-10) \approx 0.00$ to $\exp(10) \approx 22{,}000$. As s^2 becomes large (which implies clear evidence in favor of H_1), the FBF goes to 1. This can be explained by the fact that as s^2 increases, the unconstrained posterior in Equation (13) as well as the unconstrained automatic prior in Equation (12) become completely located in the constrained space of H_1. For example, in Figure 2a it can be seen that a large portion of an isodensity surface of the automatic prior for $s^2 = 9$ and $n_j = 20$ is located in the inequality constrained space $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$ (marked with thick lines). The automatic prior probability that the inequality constraints hold is equal to 0.38 in this case. As s^2 increases, both the posterior and the prior probability that the inequality constraints hold go to 1 because the posterior and the automatic prior become completely located in the inequality constrained space. Therefore, the ratio of the two probabilities in Equation (11) also goes to 1. Thus, in the FBF approach the parsimonious order constrained hypothesis that is strongly supported by the data does not receive

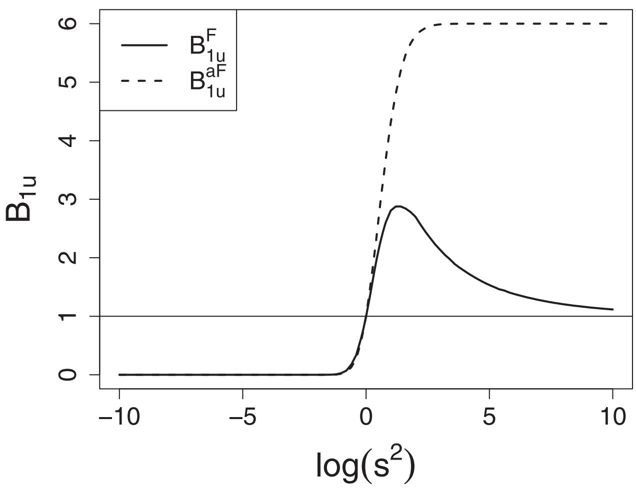

Figure 1. Fractional Bayes factor B^F_{1u} (solid line) and adjusted fractional Bayes factor B^{aF}_{1u} (dashed line) for testing $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against H_u : σ_1^2, σ_2^2, σ_3^2. The Bayes factors are plotted as a function of the sample variances $(s_1^2, s_2^2, s_3^2)' = (1, s, s^2)'$, where $s^2 \in [\exp(-10), \exp(10)]$, and for equal sample sizes of $n_1 = n_2 = n_3 = 20$.

stronger support than the more complex unconstrained hypothesis. This implies that the FBF does not function as Occam's razor in this situation. This undesirable property is a direct consequence of the fact that the automatic prior for the group variances is concentrated around the sample variances. For this reason we propose an adjustment of the FBF that corrects for this undesirable behavior when testing inequality constrained hypotheses on variances.

Adjusted Fractional Bayes Factors

In this section we present two novel extensions of the FBF approach for testing hypotheses with equality and inequality constraints on variances. The resulting criterion will be referred to as the adjusted fractional Bayes factor (aFBF).

In the aFBF the marginal likelihood is defined as

$$m_t^{aF}(\mathbf{x}, \mathbf{b}) = \frac{\int_{\Omega_t} \int_{\mathbb{R}^J} f_u(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t^a} \int_{\mathbb{R}^J} f_u(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2)^{\mathbf{b}} \pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}. \quad (15)$$

This expression has three important differences in comparison to the marginal likelihood in the FBF approach given in Equation (7). First, in the denominator in Equation (15) we integrate over an adjusted parameter space, which is denoted by Ω_t^a. The adjusted parameter space contains the same constraints as the unadjusted space Ω_t, except that each variance σ_j^2 is multiplied by a tuning parameter a_j. These tuning parameters are chosen such the automatic prior probability that the inequality constraints hold is based on prior distributions for the variances with equal scale hyperparameters (unlike in the FBF, as was observed in Equation 12). Details on the choice of the tuning parameters will be discussed below. This adjustment results in a criterion that always incorporates the parsimony of a hypothesis with inequality constraints on the variances (Böing-Messing & Mulder, 2016; Mulder, 2014b).

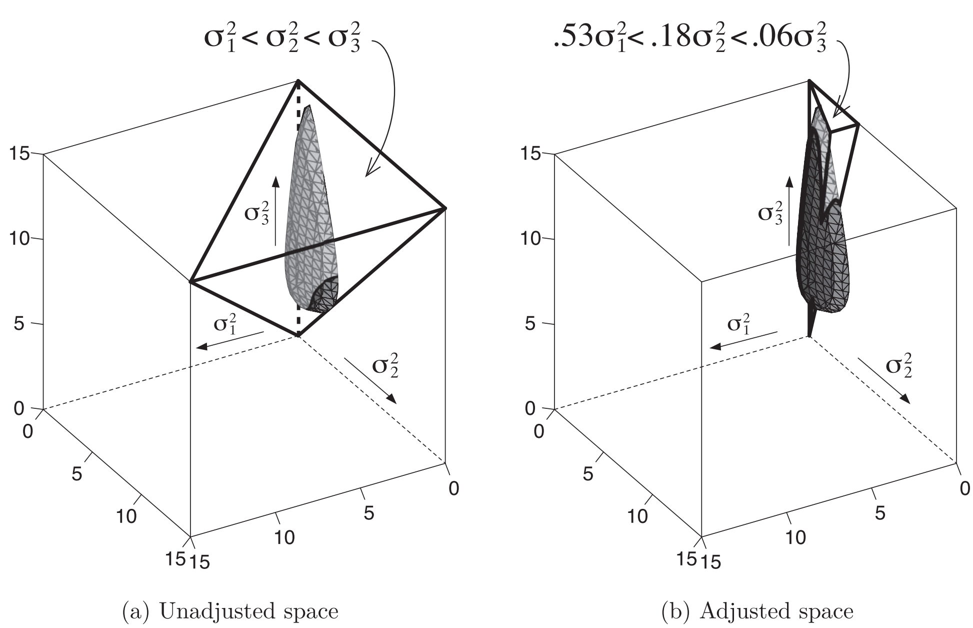

Figure 2. An isodensity surface of the automatic prior in Equation (12) for sample variances of $(s_1^2, s_2^2, s_3^2)' = (1, 3, 9)'$ and sample sizes of $n_1 = n_2 = n_3 = 20$. In Figure (a) the unadjusted parameter subspace satisfying $\sigma_1^2 < \sigma_2^2 < \sigma_3^2$ is marked with thick lines. The automatic prior probability that the inequality constraints hold equals $P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 | \mathbf{x}^b) = 0.38$. Figure (b) shows the adjusted parameter subspace satisfying $0.53\sigma_1^2 < 0.18\sigma_2^2 < 0.06\sigma_3^2$ (marked with thick lines). The adjusted automatic prior probability that the inequality constraints hold equals $P(0.53\sigma_1^2 < 0.18\sigma_2^2 < 0.06\sigma_3^2 | \mathbf{x}^b) = 1/6$.

Second, in the denominator in Equation (15) the fraction of the likelihood is based on group-specific fractions $\mathbf{b} = (b_1, \ldots, b_J)'$, where the fraction of the likelihood of group j depends on the group size according to $b_j = 2/n_j$, for $j = 1, \ldots, J$. This generalization ensures that the minimal amount of information based on two observations per group is used for automatic prior specification. This was suggested by Berger and Pericchi (2001) and De Santis and Spezzaferri (2001) for testing equality constraints on group means. Here we extend the idea to testing equality and inequality constrained hypotheses on variances. Finally, it is important to note that in the denominator in Equation (15) the likelihood and noninformative improper prior under the unconstrained hypothesis, f_u and π_u^N, are used instead of the likelihood and prior under the constrained hypothesis, f_t and π_t^N. This ensures that we integrate over the complete adjusted parameter space Ω_t^a in the denominator. For completeness, the unconstrained likelihood and prior are also used in the numerator of the marginal likelihood in the aFBF approach in Equation (15).

After some algebra (see Appendix B for a proof) the marginal likelihood in the aFBF can be expressed as

$$m_t^{aF}(\mathbf{x}, \mathbf{b}) = \widetilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) \frac{P(\boldsymbol{\sigma}^2 \in \Omega_t \mid \mathbf{x})}{P(\boldsymbol{\sigma}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}})}, \tag{16}$$

where

$$\begin{aligned}\widetilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) = & \left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}(1-b_{k_j})n_{k_j}}{2}} \\ & \prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}\right)^{-1} \\ & \left(\sum_{j=1}^{J_k}(n_{k_j}-1)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}}\left(\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}}.\end{aligned} \tag{17}$$

In Equation (17) the expression $\Gamma(\cdot)$ is the gamma function, K denotes the number of unique variances, and J_k denotes the number of groups sharing the unique variance σ_k^2, for $k = 1, \ldots, K$. Furthermore, b_{k_j} and n_{k_j} are the fraction and the sample size of the jth group sharing the unique variance σ_k^2, for $j = 1, \ldots, J_k$. In Equation (16) the adjusted parameter space Ω_t^a is defined by

$$\Omega_t^a = \{\boldsymbol{\sigma}^2 \mid (a_1\sigma_1^2, \ldots, a_K\sigma_K^2)' \in \Omega_t\}, \tag{18}$$

where the tuning parameters a_k are given by

$$a_k = \frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right) - J_k}{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j} - 1\right)s_{k_j}^2}, \tag{19}$$

for $k = 1, \ldots, K$. Furthermore, the expressions $P(\boldsymbol{\sigma}^2 \in \Omega_t \mid \mathbf{x})$ and $P(\boldsymbol{\sigma}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}})$ are the posterior and the adjusted automatic prior probability that the inequality constraints on the variances hold, respectively. These can be computed by drawing a large sample of, say, $S = 100{,}000$ draws from the unconstrained posterior and automatic prior distribution of the variances given by

$$\pi_u(\boldsymbol{\sigma}^2 \mid \mathbf{x}) = \prod_{k=1}^{K} \text{Inv-}\chi^2\left(\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k}\right) \tag{20}$$

and

$$\pi_u(\boldsymbol{\sigma}^2 \mid \mathbf{x}^{\mathbf{b}}) = \prod_{k=1}^{K} \text{Inv-}\chi^2\left(\sigma_k^2 \,\middle|\, \left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right) - J_k, \frac{\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right) - J_k}\right), \tag{21}$$

respectively. The posterior probability that the inequality constraints hold is then given by the proportion of posterior draws that satisfy the constraints, that is,

$$P(\boldsymbol{\sigma}^2 \in \Omega_t \mid \mathbf{x}) \approx \frac{1}{S}\sum_{s=1}^{S} I_{\Omega_t}(\boldsymbol{\sigma}_{post}^{2(s)}), \tag{22}$$

where $\boldsymbol{\sigma}_{post}^{2(s)}$ is the sth draw from the posterior in Equation (20), for $s = 1, \ldots, S$. Similarly, the adjusted prior probability that the inequality constraints hold is given by

$$P(\boldsymbol{\sigma}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}}) \approx \frac{1}{S}\sum_{s=1}^{S} I_{\Omega_t^a}(\boldsymbol{\sigma}_{prior}^{2(s)}), \tag{23}$$

where $\boldsymbol{\sigma}_{prior}^{2(s)}$ is the sth draw from the prior in Equation (21), for $s = 1, \ldots, S$.

Finally, it is important to note that the aFBF is scale invariant, that is, it does not depend on the scale of the outcome variable (a proof is given in Appendix C). Note that scale invariance is of crucial importance because in comparing educational performances in different grades, for example, it should not matter whether students' performances are rated on a scale from 0 to 10 or from 0 to 100.

Adjusted Fractional Bayes Factors for an Inequality Constrained Test

Now we apply the aFBF to the test of $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ against $H_u : \sigma_1^2, \sigma_2^2, \sigma_3^2$. As noted above, the adjusted parameter space contains the same constraints as the unadjusted space, except that the variances are multiplied by tuning parameters a_j which correct for the differences between the observed sample variances. Thus, the adjusted parameter space under H_1 is given by $\Omega_1^a = \{\boldsymbol{\sigma}^2 \mid a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2\}$, with $a_j = n_j/(2(n_j - 1)s_j^2)$. Furthermore, the fractions are given by $\mathbf{b} = (b_1, b_2, b_3)'$, with $b_j = 2/n_j$, for $j = 1$, 2 and 3. The aFBF for H_1 against H_u can then be written as

$$B_{1u}^{aF} = \frac{P(\boldsymbol{\sigma}^2 \in \Omega_1 \mid \mathbf{x})}{P(\boldsymbol{\sigma}^2 \in \Omega_1^a \mid \mathbf{x}^{\mathbf{b}})} = \frac{P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2 \mid \mathbf{x})}{P(a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2 \mid \mathbf{x}^{\mathbf{b}})}. \tag{24}$$

Note that the posterior probability in the numerator is identical to that in the FBF in Equation (11). On the other hand, the automatic prior probability of the adjusted ordering in the denominator is computed using the automatic prior distribution

$$\pi_u(\boldsymbol{\sigma}^2|\mathbf{x^b}) = \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2 \middle| b_j n_j - 1, \frac{b_j(n_j-1)s_j^2}{b_j n_j - 1}\right)$$
$$= \prod_{j=1}^{3} \text{Inv-}\chi^2\left(\sigma_j^2 \middle| 1, \frac{2(n_j-1)s_j^2}{n_j}\right). \quad (25)$$

First note that the prior degrees of freedom are equal to 1, which implies minimal information for any group size n_j. Regarding the scale hyperparameter, standard mathematical statistics dictates that multiplying a random variable having a scaled inverse-χ^2 distribution by a constant, say a, results in a new random variable having a scaled inverse-χ^2 distribution where the original scale parameter is multiplied by a. For this reason, because $\sigma_j^2|\mathbf{x^b} \sim$ Inv-$\chi^2(1, 2(n_j - 1)s_j^2/n_j)$, it automatically holds that

$$a_j\sigma_j^2|\mathbf{x^b} \sim \text{Inv-}\chi^2\left(1, a_j\frac{2(n_j-1)s_j^2}{n_j}\right) = \text{Inv-}\chi^2(1,1), \quad (26)$$

for $j = 1$, 2, and 3. Thus, the multiplication by the tuning parameters results in equal automatic prior distributions for $a_1\sigma_1^2$, $a_2\sigma_2^2$, and $a_3\sigma_3^2$. Because these distributions are equal, all six possible adjusted orderings "$a_1\sigma_1^2 < a_2\sigma_2^2 < a_3\sigma_3^2$", . . . , "$a_3\sigma_3^2 < a_2\sigma_2^2 < a_1\sigma_1^2$" are equally likely under the automatic prior. Therefore, the automatic prior probability of each adjusted ordering is equal to 1/6. Consequently, the Bayes factor in Equation (24) is equal to

$$B_{1u}^{aF} = 6 \times P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x}). \quad (27)$$

Again, we consider data with $n_j = n = 20$ observations in each group with sample variances of $s_1^2 = 1$, $s_2^2 = s$, and $s_3^2 = s^2$, and we compute the aFBF for H_1 against H_u while letting s^2 increase from $\exp(-10) \approx 0$ to $\exp(10) \approx 22{,}000$. The results are shown in Figure 1. It can be seen that the aFBF (dashed line) converges to 6 as s^2 increases. This is a result of the posterior probability in Equation (27), which goes to 1 as s^2 increases, similar as in the FBF. Unlike in the FBF, however, the prior probability of the adjusted ordering is equal to 1/6. To give some more intuition, Figure 2b displays the adjusted parameter space when the sample variances are equal to $s_1^2 = 1$, $s_2^2 = 3$, and $s_3^2 = 9$, and the group sizes are equal to $n_j = n = 20$, for $j = 1$, 2, and 3. The plot illustrates how the adjusted parameter space adapts to the observed sample variances to ensure that the automatic prior probability of the adjusted ordering always equals 1/6. Because the aFBF for H_1 against H_u converges to 6, it can be argued that the order constrained hypothesis H_1 is 6 times more parsimonious than the unconstrained hypothesis.

Finally, note that in practice we do not recommend testing an inequality constrained hypothesis against the unconstrained hypothesis as in the above example. The reason is that the aFBF (and Bayes factors in general) is then bounded (e.g., by 6 in the case of $J = 3$ groups). This implies that we can never get decisive evidence in favor of H_1, even when observing very large effects in the direction of H_1 with very large samples. The main reason for testing H_1 against H_u in the above example was to illustrate how the parsimony of an inequality constrained hypothesis on variances is incorporated in the FBF and the aFBF. Generally, we would recommend testing an inequality constrained hypothesis H_1 against its complement H_2 : not H_1 to avoid the issue of a bounded Bayes factor. For this test the aFBF would be equal to

$$B_{12}^{aF} = \frac{B_{1u}^{aF}}{B_{2u}^{aF}} = \frac{6 \times P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x})}{6/5 \times P(\text{not } \sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x})}$$
$$= 5 \times \frac{P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x})}{1 - P(\sigma_1^2 < \sigma_2^2 < \sigma_3^2|\mathbf{x})}, \quad (28)$$

which does not have an upper bound.

Posterior Probabilities of the Hypotheses

When there are more than two hypotheses under investigation, it is useful to transform Bayes factors to posterior probabilities of the hypotheses. Here we show how to do this when working with the aFBF. To compute the posterior probabilities we first need to specify the prior probabilities of the hypotheses, denoted by $P(H_t)$, for $t = 1, \ldots, T$, where T is the number of hypotheses that are tested. These prior probabilities quantify how plausible each hypothesis is before observing the data. After observing the data, the prior probabilities can be updated using the marginal likelihoods from the aFBF in Equation (16) as follows:

$$P^{aF}(H_t|\mathbf{x},\mathbf{b}) = \frac{m_t^{aF}(\mathbf{x},\mathbf{b})P(H_t)}{\sum_{t'=1}^{T} m_{t'}^{aF}(\mathbf{x},\mathbf{b})P(H_{t'})}. \quad (29)$$

The resulting posterior probabilities $P^{aF}(H_t|\mathbf{x},\mathbf{b})$ quantify how plausible each hypothesis is after observing the data, for $t = 1, \ldots, T$. Note that the superscript aF is added to make it explicit that the posterior probabilities are computed using the marginal likelihoods based on the aFBF approach (see Equation 16).

The default (or objective) choice in the literature is to set equal prior probabilities for the hypotheses, that is, $P(H_1) = \cdots = P(H_T) = 1/T$, which implies that it is assumed that all hypotheses are equally likely a priori (e.g., Berger & Mortera, 1999; Hoijtink, 2011; Mulder, Hoijtink, & de Leeuw, 2012). A consequence is that the ratio of the posterior probabilities of a pair of hypotheses is equal to the respective Bayes factor of these hypotheses. Because the prior probabilities sum to 1 (as well as the posterior probabilities), it is implicitly assumed that the true hypothesis is present in the set of constrained hypotheses under investigation. To ensure that this is the case it is recommended to always include the complement hypothesis when testing a set of constrained hypothesis on the variances. This was also done in the Math Garden example by including the complement hypothesis H_3 in Equation (4). Note that it is not recommended to set the prior probability of a hypothesis equal to the proportion of the unconstrained parameter space that it covers (e.g., 1/6 for $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$). In that case the posterior probability of an inequality constrained hypothesis does not properly take the parsimony due to the inequality constraints into account (for details see Mulder, 2014a). Furthermore, the proportion of the unconstrained parameter space that is covered by a hypothesis involving at least one equality constraint is 0 (e.g., $H_2 : \sigma_1^2 = \sigma_2^2 < \sigma_3^2$ describes a plane in the unconstrained space, which has a volume of 0). However, a prior probability of 0 results in a posterior probability of 0 (see Equation 29), which means that there can never be evidence in favor of an equality constrained hypothesis.

Simulation Study: Performance of the Adjusted Fractional Bayes Factor

The goal of our simulation study is to assess the performance of the adjusted fractional Bayes factor when testing equality and inequality constrained hypotheses on variances. Our focus is both on consistency (i.e., does the aFBF select the true hypothesis when the sample size is large) and small-sample performance.

Design

The performance of the aFBF is examined as a function of the following four factors:

1. *Number of groups:* We compared variances of $J = 3$ and $J = 5$ groups.

2. *Population:* For each of the two numbers of groups we considered five populations differing in the structure of the population variances. An overview is given in Table 2. The first was a null population in which all population variances were equal, $\sigma_1^2 = \cdots = \sigma_J^2$. The second population was one in which the variances followed the hypothesized order $\sigma_1^2 < \cdots < \sigma_J^2$. We refer to this population as the order population. The mixed population featured equalities as well as inequalities among the variances. For $J = 3$ groups the structure of the population variances was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2$, whereas for $J = 5$ groups it was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$. The near order population was identical to the order population with the exception that the order of the two groups with the largest variances was reversed, $\sigma_1^2 < \cdots < \sigma_J^2 < \sigma_{J-1}^2$. Finally, in the reverse order population the order of the population variances was the opposite of that in the order population, $\sigma_J^2 < \cdots < \sigma_1^2$. Note that the reverse order is maximally different from the hypothesized order. We included the near order and the reverse order population to check how much data is needed to detect that the hypothesized order is slightly different from the true order (near order population) or very different from the true order (reverse order population).

3. *Effect size:* In all populations except the null population we considered three effect sizes: small, medium, and large. The effect size is given by the ratio of the largest population variance to the smallest population variance. To our knowledge no guidelines exist as to what population variance ratios constitute a small, medium, and large effect. We therefore based our effect sizes on well-known guidelines for testing equality of means of two independent populations. These guidelines state that the power to detect a small, medium, and large effect equals 0.8 for $\alpha = .05$ and sample sizes of 310, 51, and 21 in each group, respectively (Faul, Erdfelder, Buchner, & Lang, 2009). We used these numbers to determine the population variances in our simulation study in four steps: First, we used the sample sizes of 310, 51, 21 to determine the noncentrality parameter λ of the noncentral F-distribution such that the power for testing equality of variances of two independent populations equals 0.8. For a small, medium, and large effect, we obtain values of λ of 100.74, 49.94, and 38.80, respectively. Second, we computed the population variance ratio as $\text{VR} = (n - 1 + \lambda)/(n - 1)$, which equals the expected value of the noncentral F-distribution. Here the common sample size n equals 310, 51, and 21 if λ equals 100.74, 49.94, and 38.80, respectively. The resulting ratios are 1.33, 2.00, and 2.94 for a small, medium, and large effect, respectively. Third, to determine σ_J^2/σ_1^2 for $J = 3$ and 5 groups, we computed the $(J - 1)/J$ quantile of a uniform distribution with minimum value 1 and maximum value $2 \times \text{VR} - 1$. This results in population variance ratios that increase with the number of groups J, which is supported by empirical findings (see, e.g., Ruscio & Roche, 2012). In all populations we set $\sigma_1^2 = 1$, so that σ_J^2 is determined by the population variance ratio. Fourth, we computed the intermediate population variances as $\sigma_j^2 = (\sigma_J^2)^{(j-1)/(J-1)}$ for $j = 2, \ldots, J - 1$. As a result, the ratio of adjacent population variances is constant, that is, $\sigma_2^2/\sigma_1^2 = \cdots = \sigma_J^2/\sigma_{J-1}^2$. Table 3 gives an overview of all population variances used in the simulation study. Note that in the mixed population with $J = 3$ groups we used the population variance ratios from the $J = 2$ groups case, that is, 1.33, 2.00, and 2.94. We did so because, in fact, there are only two distinct variances in this population (cf. Table 2). Similarly, in the mixed population with $J = 5$ groups we used the population variance ratios from the $J = 3$ groups case.

4. *Sample size:* We used a balanced design with common sample sizes of 5, 10, 20, 50, 100, 200, 500, 1,000, 2,000, and 5,000.

Thus, in total there were 260 conditions, 2 (number of groups) $\times$ 10 (sample size) = 20 for the null population and 2 (number of groups) $\times$ 4 (population) $\times$ 3 (effect size) $\times$ 10 (sample size) = 240 for the remaining four populations.

Table 2
Structure of Population Variances in Five Populations for $J \in \{3, 5\}$ Groups

Population	$J = 3$	$J = 5$
Null	$\sigma_1^2 = \sigma_2^2 = \sigma_3^2$	$\sigma_1^2 = \cdots = \sigma_5^2$
Order	$\sigma_1^2 < \sigma_2^2 < \sigma_3^2$	$\sigma_1^2 < \cdots < \sigma_5^2$
Mixed	$\sigma_1^2 < \sigma_2^2 = \sigma_3^2$	$\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$
Near order	$\sigma_1^2 < \sigma_3^2 < \sigma_2^2$	$\sigma_1^2 < \cdots < \sigma_5^2 < \sigma_4^2$
Reverse order	$\sigma_3^2 < \sigma_2^2 < \sigma_1^2$	$\sigma_5^2 < \cdots < \sigma_1^2$

Hypotheses and Data Generation

In each of the five populations we tested three hypotheses. An overview is given in Table 4. In the null population, the order population, the near order population, and the reverse order population we tested the following three hypotheses: $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$, $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$, and H_2 : not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that H_2 : not $\sigma_1^2 < \cdots < \sigma_J^2$ is equivalent to H_2 : not (H_0 or H_1) because the probability of the event that the variances are exactly equal is 0 under the unconstrained hypothesis. In Table 4 the true hypoth-

Table 3
Population Variances in the Simulation Study

Population	Effect	$J = 3$ σ_1^2	σ_2^2	σ_3^2	$J = 5$ σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2
Null	No	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Order	Small	1.00	1.20	1.43	1.00	1.11	1.23	1.37	1.52
	Medium	1.00	1.53	2.33	1.00	1.27	1.61	2.05	2.60
	Large	1.00	1.89	3.59	1.00	1.42	2.03	2.88	4.10
Mixed	Small	1.00	1.33	1.33	1.00	1.20	1.20	1.43	1.43
	Medium	1.00	2.00	2.00	1.00	1.53	1.53	2.33	2.33
	Large	1.00	2.94	2.94	1.00	1.89	1.89	3.59	3.59
Near order	Small	1.00	1.43	1.20	1.00	1.11	1.23	1.52	1.37
	Medium	1.00	2.33	1.53	1.00	1.27	1.61	2.60	2.05
	Large	1.00	3.59	1.89	1.00	1.42	2.03	4.10	2.88
Reverse order	Small	1.43	1.20	1.00	1.52	1.37	1.23	1.11	1.00
	Medium	2.33	1.53	1.00	2.60	2.05	1.61	1.27	1.00
	Large	3.59	1.89	1.00	4.10	2.88	2.03	1.42	1.00

esis (i.e., the hypothesis that correctly describes the structure of the population variances) is flagged with an asterisk (H^*). Note that for the near order and the reverse order population the true hypothesis is contained in the complement H_2. In the mixed population with $J = 3$ groups we tested $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$, $H_1 : \sigma_1^2 < \sigma_2^2 = \sigma_3^2$, and $H_2 : \sigma_1^2 < (\sigma_2^2, \sigma_3^2)$. Here H_2 states that the variances in Groups 2 and 3 are larger than in Group 1, but not necessarily equal. In the mixed population with $J = 5$ groups we tested the corresponding hypotheses $H_0 : \sigma_1^2 = \cdots = \sigma_5^2$, $H_1 : \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$, and $H_2 : \sigma_1^2 < (\sigma_2^2, \sigma_3^2) < (\sigma_4^2, \sigma_5^2)$.

In each of the 260 conditions we generated 1,000 data sets. The population variances were specified according to Table 3. In all conditions we set $\mu_1 = \cdots = \mu_J = 0$. We may do so because the aFBF is independent of the population means (in Equations 16 and 17 it can be seen that the marginal likelihood in the aFBF approach does not depend on the sample means). For each of the 1,000 data sets we computed the evidence in favor of the true hypothesis. We used two measures of evidence: The first is the logarithm of the Bayes factor in favor of the true hypothesis H_t, $\log(B_{tt'}^{aF})$. The second measure is the posterior probability of the true hypothesis, $P^{aF}(H_t \mid \mathbf{x}, \mathbf{b})$, which was computed assuming equal prior probabilities of the hypotheses. The log Bayes factors and posterior probabilities were computed using minimal fractions of $b_j = 2/n_j$, for $j = 1, \ldots, J$. Eventually, we computed the median of the 1,000 log Bayes factors and posterior probabilities.

Results

The results of the simulation study are shown in Figures 3 to 7. Each figure shows the results for one of the five populations we considered. The plots show the median log Bayes factors in favor of the true hypothesis (left-hand column) and the median posterior probability of the true hypothesis (flagged with an asterisk in Table 4; right-hand column) for $J = 3$ groups (top row) and $J = 5$ groups (bottom row) as a function of the common sample size $n_1 = \cdots = n_J = n$. For the null population the results for $J = 3$ and $J = 5$ groups are combined in one pair of plots, see Figure 3. Two important general conclusions can be drawn from the figures. First, the aFBF is consistent. For all numbers of groups, populations, and effect sizes the posterior probability of the true hypothesis was equal or close to 1 for a common sample size of 5,000. Second, the performance of the aFBF was similar for $J = 3$ and $J = 5$ groups, with the relevant differences being that for $J = 5$ groups the null hypothesis received stronger support and larger sample sizes were needed to reject a false null hypothesis. We now focus on small-sample performance of the aFBF for each population separately.

Null population. Figure 3 shows the simulation results for the null population. The plots show that the evidence in favor of the true hypothesis H_0 increased with sample size. The log Bayes factor $\log(B_{01}^{aF})$ was consistently larger than $\log(B_{02}^{aF})$ because under

Table 4
Hypotheses Tested in the Simulation Study

Population	Tested hypotheses		
Null	$H_0^* : \sigma_1^2 = \cdots = \sigma_J^2$	$H_1 : \sigma_1^2 < \cdots < \sigma_J^2$	$H_2 :$ not $\sigma_1^2 < \cdots < \sigma_J^2$
Order	$H_0 : \sigma_1^2 = \cdots = \sigma_J^2$	$H_1^* : \sigma_1^2 < \cdots < \sigma_J^2$	$H_2 :$ not $\sigma_1^2 < \cdots < \sigma_J^2$
Mixed, $J = 3$	$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$	$H_1^* : \sigma_1^2 < \sigma_2^2 = \sigma_3^2$	$H_2 : \sigma_1^2 < (\sigma_2^2, \sigma_3^2)$
Mixed, $J = 5$	$H_0 : \sigma_1^2 = \cdots = \sigma_5^2$	$H_1^* : \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$	$H_2 : \sigma_1^2 < (\sigma_2^2, \sigma_3^2) < (\sigma_4^2, \sigma_5^2)$
Near order	$H_0 : \sigma_1^2 = \cdots = \sigma_J^2$	$H_1 : \sigma_1^2 < \cdots < \sigma_J^2$	$H_2^* :$ not $\sigma_1^2 < \cdots < \sigma_J^2$
Reverse order	$H_0 : \sigma_1^2 = \cdots = \sigma_J^2$	$H_1 : \sigma_1^2 < \cdots < \sigma_J^2$	$H_2^* :$ not $\sigma_1^2 < \cdots < \sigma_J^2$

Note. In each population we tested three hypotheses. The true hypothesis is flagged with an asterisk (H^*). Here $J \in \{3, 5\}$ indicates the number of groups.

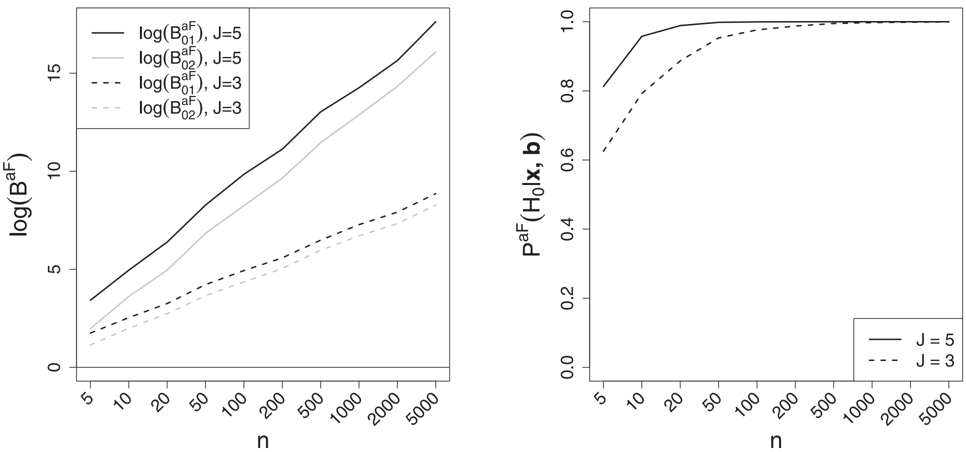

Figure 3. Simulation results for a null population in which all population variances were equal, $\sigma_1^2 = \cdots = \sigma_J^2$, for $J = 3$ groups (dashed lines) and $J = 5$ groups (solid lines). We tested the true hypothesis $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$ against the two competing hypotheses $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ and H_2 : not $\sigma_1^2 < \cdots < \sigma_J^2$. The plots show the median log Bayes factors (left-hand plot) testing H_0 against H_1 (black lines) and H_0 against H_2 (gray lines) and the median posterior probability of H_0 (right-hand plot) as a function of the common sample size $n_1 = \cdots = n_J = n$.

H_0 the order constrained hypotheses H_1 fits worse than the complement H_2. This is because H_1 is more restrictive than H_2. In the right-hand plot we see that for samples of size $n = 5$ the posterior probability of H_0 was greater than 0.6, and samples as small as $n = 10$ yielded a posterior probability of about 0.8. The probability is so high even for small samples because neither H_1 nor H_2 are good competitors to H_0, particularly so for $J = 5$ groups.

Order population. Figure 4 shows the simulation results for the order population. The plots illustrate that the evidence in favor of the true hypothesis H_1 did not increase with sample size when the effect was small. This is a consequence of the fact that small effects can be better explained by the null hypothesis than by the order constrained hypothesis when the sample size is small. The posterior probability of the true hypothesis H_1 was at least 0.8 for sample sizes of about 500 (small effect), 100 (medium effect), and 50 (large effect), respectively. Finally, note that the gray lines for the log Bayes factor $\log(B_{12}^{aF})$ are discontinued at some point. This is due to numerical reasons: In the computation of the discontinued log Bayes factors we had to divide by the posterior probability that the inequality constraints do not hold. This was estimated by the proportion of draws from the unconstrained posterior distribution for which these constraints do not hold. For large samples this proportion was often 0, so that the corresponding log Bayes factor was undefined. If this happened for the majority of the 1,000 replications in the simulation, then the median log Bayes factor was undefined as well. Note that theoretically the discontinued log Bayes factors keep increasing because the posterior probability that the inequality constraints do not hold approaches 0 as the sample size increases.

Mixed population. The results for the mixed population are shown in Figure 5. Similar to the order population, the evidence in favor of the true hypothesis did not increase with sample size when the effect was small. For $J = 3$ groups and a small effect the evidence only increased for sample sizes larger than 50. Actually, in this case the log Bayes factor $\log(B_{10}^{aF})$ favored the null hypothesis until the sample size surpassed the $n = 200$ mark. The reason is the same as for the order population, namely, that small effects can be better explained by the null hypothesis when the sample size is small. The posterior probability was above 0.8 for samples of size 500 (small effect), 50 to 100 (medium effect), and 50 (large effect), respectively. Note that it approached 1 somewhat more slowly than in the order population. This is due to the similarity of H_1 and H_2. Finally, note that the log Bayes factor $\log(B_{12}^{aF})$ did not depend on the effect size. It was approximately the same under all effects, which can be seen from the three gray lines overlapping. This is because H_1 and H_2 essentially state the same effect, namely, that σ_1^2 is smaller than σ_2^2 and σ_3^2.

Near order population. Figure 6 shows the simulation results for the near order population. Again, the evidence in favor of the true hypothesis did not generally increase with sample size. For a small effect it only increased for sample sizes larger than about 100. For $J = 3$ groups the posterior probability of the true hypothesis reached values of at least 0.8 for samples of size 1,000 (small effect), 100 (medium effect), and 50 (large effect), respectively. For $J = 5$ groups substantially larger samples were required (5,000, 1,000, and 200, respectively). This is mainly because the ratio of adjacent population variances is smaller in the $J = 5$ groups case (see Table 3), which makes it more difficult to detect that the two largest population variances are ordered as $\sigma_J^2 < \sigma_{J-1}^2$ instead of $\sigma_{J-1}^2 < \sigma_J^2$. Similar to the order population, the log Bayes factor $\log(B_{21}^{aF})$ (gray lines) could not be computed for larger sample sizes due to numerical reasons.

Reverse order population. The evidence in favor of the true hypothesis did not generally increase with sample size, see Figure 7. For instance, for a small effect and $J = 5$ groups the evidence only increased for sample sizes larger than 200. The evidence in favor of the true hypothesis increased faster for the reverse order than for the near order population because the reverse order population is less in agreement with the order constrained hypothesis H_1 than the near order population. The posterior probability of

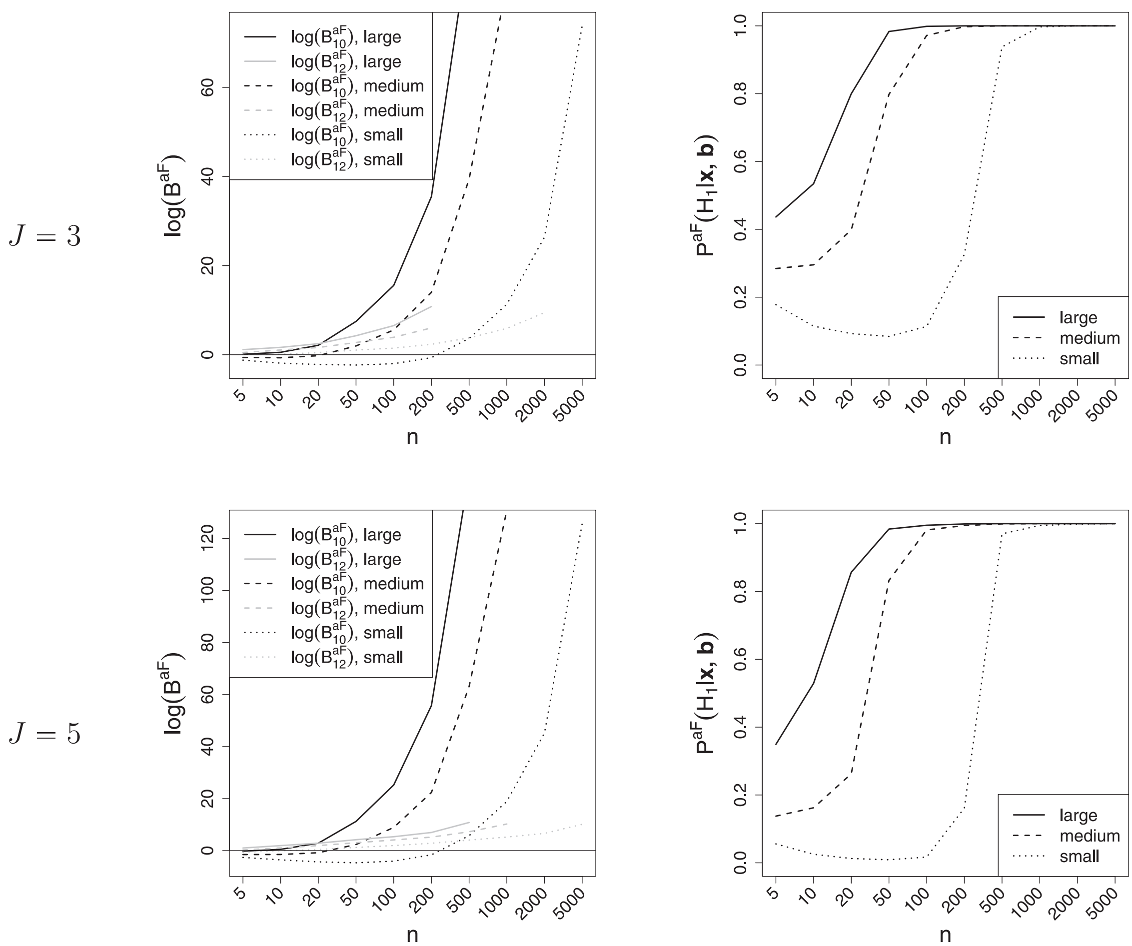

Figure 4. Simulation results for an order population in which the structure of the population variances was $\sigma_1^2 < \cdots < \sigma_J^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested the true hypothesis $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$ against the two competing hypotheses $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$ and H_2 : not $\sigma_1^2 < \cdots < \sigma_J^2$. The plots show the median log Bayes factors (left-hand column) testing H_1 against H_0 (black lines) and H_1 against H_2 (gray lines) and the median posterior probability of H_1 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

the true hypothesis was greater than 0.8 for sample sizes of 500 (small effect), 100 (medium effect), and 50 (large effect), respectively. Again, we see discontinued log Bayes factors due to numerical reasons.

Conclusion

In conclusion, the results of the simulation show that the aFBF performed well in all five populations we considered. In particular, the results indicate that the aFBF is consistent in the sense that it selects the true hypothesis if the sample size is large enough. Naturally, for small effects we needed larger samples to detect the true hypothesis than for large effects.

We also performed the simulation with unequal group sizes to check for robustness of the results obtained with equal group sizes. All settings except the sample sizes were identical to the simulation with equal group sizes. We provide the sample sizes and results of the simulation with unequal group sizes in the supplemental material to this article. The results confirm the findings from the simulation with equal group sizes discussed above.

Illustrative Example: The Math Garden (Continued)

After logging into the Math Garden, children are directed to a page showing a garden in which plants represent games covering different domains of mathematics (see Figure 8a). In this illustrative example we focus on the four most played games: addition, subtraction, multiplication, and division. Each of these games consists of over 700 items ranging from easy (e.g., 2 + 2) to difficult (e.g., 340 + 87). Figure 8b shows an exemplary addition item. By clicking on a plant the player starts a session of 15 items. The items are adaptively selected based on a player's ability. The system takes both accuracy of responses and response times into account to estimate a player's ability. For details on the Math Garden and the underlying IRT model we refer the interested

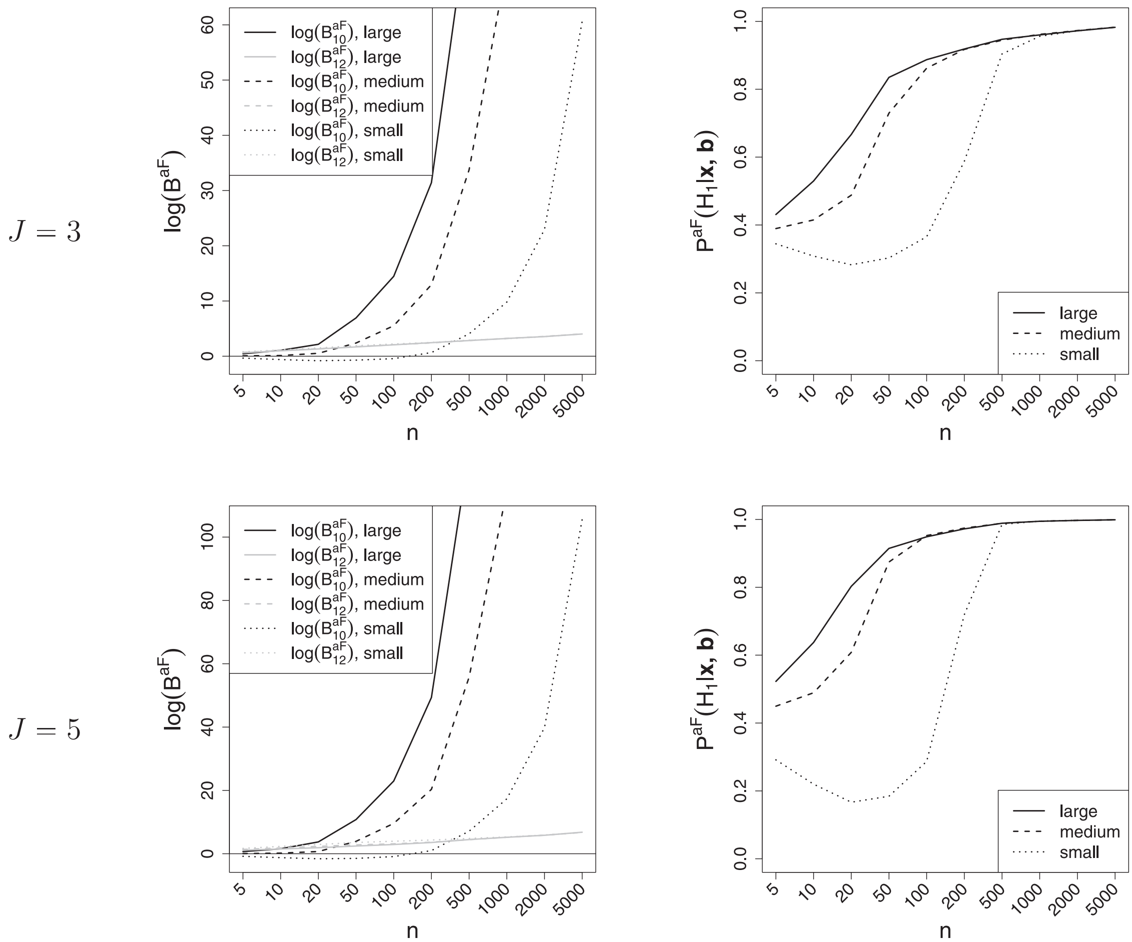

Figure 5. Simulation results for a mixed population. For $J = 3$ groups the structure of the population variances was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2$, whereas for $J = 5$ groups it was $\sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). For $J = 3$ groups we tested the true hypothesis $H_1 : \sigma_1^2 < \sigma_2^2 = \sigma_3^2$ against the two competing hypotheses $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$ and $H_2 : \sigma_1^2 < (\sigma_2^2, \sigma_3^2)$. For $J = 5$ groups we tested the true hypothesis $H_1 : \sigma_1^2 < \sigma_2^2 = \sigma_3^2 < \sigma_4^2 = \sigma_5^2$ against $H_0 : \sigma_1^2 = \cdots = \sigma_5^2$ and $H_2 : \sigma_1^2 < (\sigma_2^2, \sigma_3^2) < (\sigma_4^2, \sigma_5^2)$. The plots show the median log Bayes factors (left-hand column) testing H_1 against H_0 (black lines) and H_1 against H_2 (gray lines) and the median posterior probability of H_1 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$.

reader to Klinkenberg et al. (2011) and Maris and van der Maas (2012).

We used two criteria for extracting ability estimates from the Math Garden database. The first criterion concerns the grade a student is in. Aunola et al. (2004) hypothesize that systematic instruction at school functions as a sort of treatment that results in an increase or a decrease of the variability of abilities across grades. It thus makes sense to only consider grades in which the treatment is administered to the students. In the Netherlands children are taught addition and subtraction at school from Grade 1 through Grade 5. For the addition and the subtraction domain we therefore extracted ability estimates of students in Grades 1 through 5. Multiplication and division is taught from Grade 3 through 6, which is why for these two domains we extracted ability estimates of students in these grades. The second criterion we used is that children have to have played at least 45 items (i.e., three sessions) in the week prior to extraction. The reason for this is twofold. First, the more items a student plays the more precise their ability can be estimated. Experience has shown that after 45 items ability estimates are reasonably precise and stable. Second, we require children to have played the items in one week in order to avoid that there is too much learning going on due to treatment at school.

Table 5 shows the sample size and sample variance for each grade and mathematical domain. We use the symbols +, −, ×, and ÷ to refer to the corresponding game in the Math Garden. Furthermore, the table shows the variance ratio, which is given by the ratio of a sample variance to the smallest sample variance in the corresponding domain. In the addition and the subtraction domain it can be seen that the sample variances do not follow an increasing order. The variance decreases from Grade 1 to Grade 2, and subsequently increases from Grade 2 to Grade 5. In the multiplication and the division domain, however, the

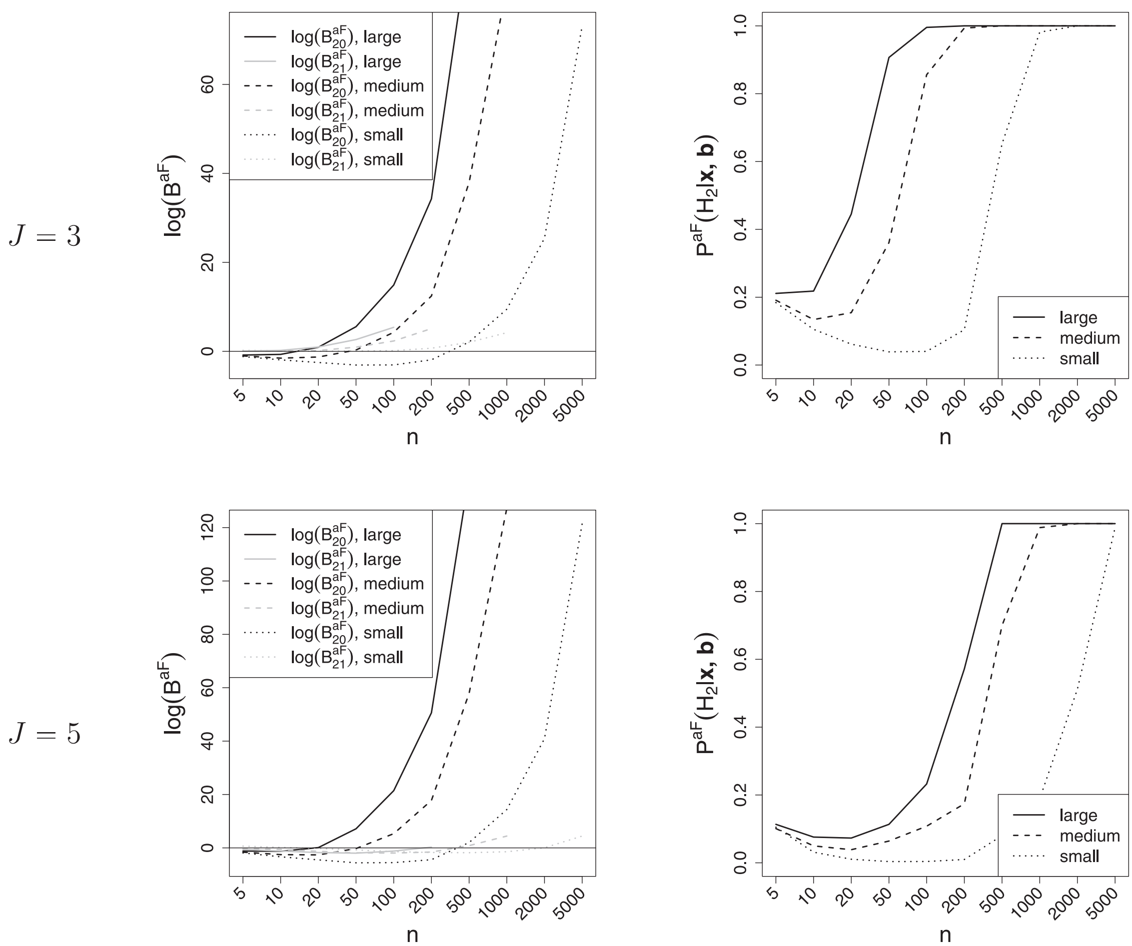

Figure 6. Simulation results for a near order population in which the structure of the population variances was $\sigma_1^2 < \cdots < \sigma_J^2 < \sigma_{J-1}^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested three hypotheses: $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$, $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$, and H_2 : not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that the true hypothesis is the complement H_2. The plots show the median log Bayes factors (left-hand column) testing H_2 against H_0 (black lines) and H_2 against H_1 (gray lines) and the median posterior probability of H_2 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

sample variances follow an increasing order from Grade 3 to Grade 6.

Table 6 shows the posterior probability of the hypotheses $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$, $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$, $H_2 : \sigma_J^2 < \cdots < \sigma_1^2$, and H_3 : not (H_0 or H_1 or H_2) for each domain. We computed the posterior probabilities assuming equal prior probabilities of the hypotheses. The posterior probabilities are (close to) 0.00 or 1.00 due to the large sample sizes in combination with the considerable effect sizes (cf. the results of the simulation study). One immediate conclusion we can draw is that there is no evidence in favor of either H_0 or H_2 in any of the domains, as can be seen from their posterior probabilities being 0.00. The hypotheses of equality of variances and decreasing variances can therefore safely be rejected. Furthermore, in the addition and the subtraction domain we can rule out H_1 given posterior probabilities of 0.00 and 0.03, respectively. The decrease in variance from Grade 1 to Grade 2 in combination with the large sample sizes makes an increasing order of the variances highly unlikely. We conclude that in the addition and the subtraction domain something other than H_0, H_1, and H_2 is going on, as is indicated by the posterior probabilities of the complement H_3 being 1.00 and 0.97, respectively. In the multiplication and the division domain, however, there is very strong evidence in favor of an increase in variance, with posterior probabilities of H_1 of 1.00. In these domains we can rule out H_0, H_2, and H_3, as is indicated by posterior probabilities of these hypotheses of 0.00.

Software Application for Computing the Adjusted Fractional Bayes Factor

We provide a Shiny application for computing the adjusted fractional Bayes factor. Shiny (Chang, Cheng, Allaire, Xie, & McPherson, 2015) is a framework for creating interactive applications using the R language for statistical computing (R Core Team, 2015). The advantage of Shiny applications is that the user does not need to read or write R code.

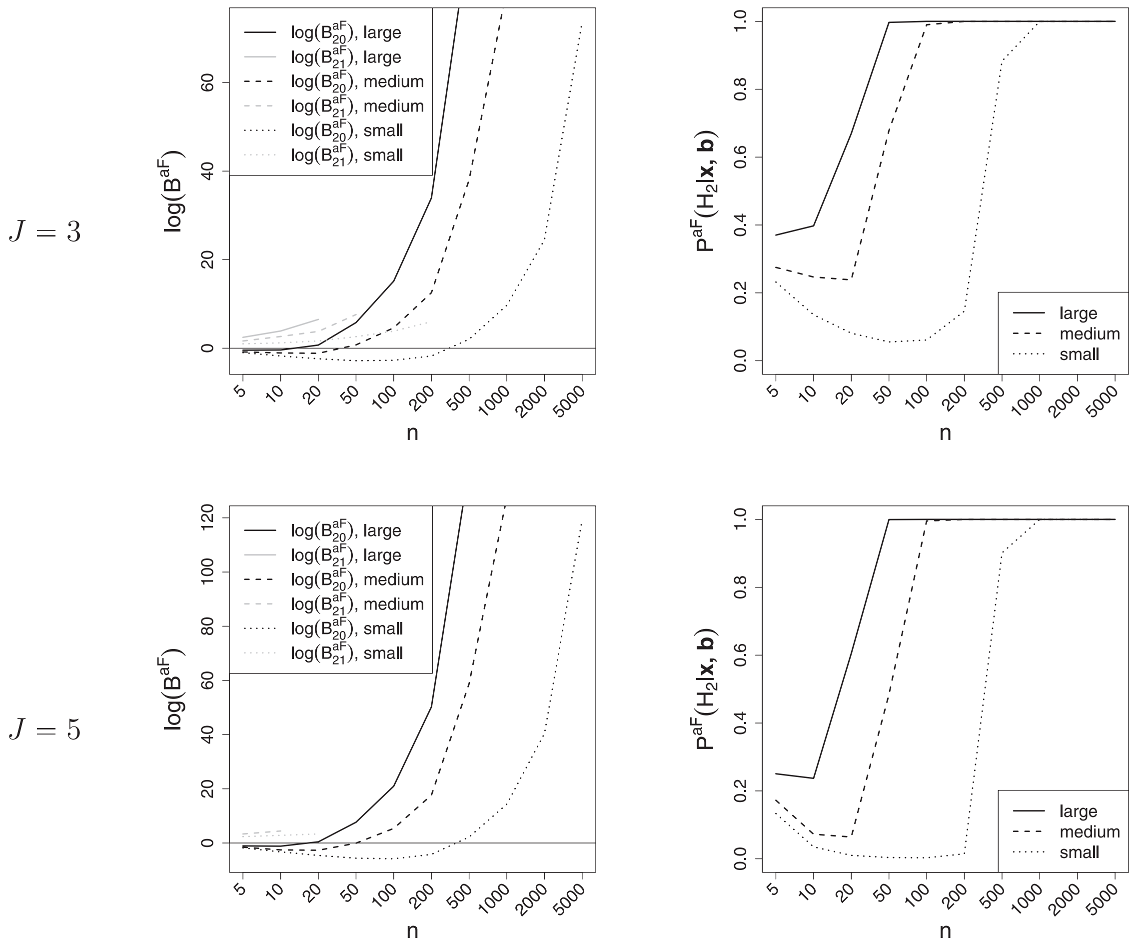

Figure 7. Simulation results for a reverse order population in which the structure of the population variances was $\sigma_J^2 < \cdots < \sigma_1^2$, with $J \in \{3, 5\}$. We considered three effect sizes: small (dotted lines), medium (dashed lines), and large (solid lines). We tested three hypotheses: $H_0 : \sigma_1^2 = \cdots = \sigma_J^2$, $H_1 : \sigma_1^2 < \cdots < \sigma_J^2$, and H_2 : not $\sigma_1^2 < \cdots < \sigma_J^2$. Note that the true hypothesis is the complement H_2. The plots show the median log Bayes factors (left-hand column) testing H_2 against H_0 (black lines) and H_2 against H_1 (gray lines) and the median posterior probability of H_2 (right-hand column) as a function of the common sample size $n_1 = \cdots = n_J = n$. In the log Bayes factors plots the gray lines are discontinued due to numerical reasons (see text).

Figure 9 shows two screenshots of our Shiny application. On the left-hand side of Figure 9a one can see the "Mandatory input" tab panel. Here the user needs to specify the sample variances, sample sizes, and hypotheses. The screenshot shows the input for the addition domain in the Math Garden example. As can be seen, the hypotheses need to be specified using group numbers $1, \ldots, J$. For example, the hypothesis $H_1 : \sigma_1^2 < \cdots < \sigma_5^2$ from the Math Garden example is specified as "1<2<3<4<5". Note that inequality constraints need to be specified using the less-than symbol (<); the greater-than symbol (>) is not supported. The complement of an order constrained hypothesis can be specified using the string "not" in the beginning (e.g., "not 1<2<3<4<5"). Note that the complement of a hypothesis containing at least one equality constraint is equivalent to the unconstrained hypothesis. This is because the probability of the event that two or more variances are exactly equal is 0 under the unconstrained hypothesis. For example, the hypothesis $H_1 : \sigma_1^2 = \sigma_2^2 < \sigma_3^2$ describes a plane in the unconstrained space, which has a probability of 0 (in the sense that the volume is 0). The complement H_2 : not $\sigma_1^2 = \sigma_2^2 < \sigma_3^2$ comprises the entire space except the plane in H_1, which is mathematically equivalent to the unconstrained space. Hence, the complement H_2 is equivalent to the unconstrained hypothesis H_u : $\sigma_1^2, \sigma_2^2, \sigma_3^2$. For the same reason the hypothesis H_3 : not ($\sigma_1^2 = \cdots = \sigma_5^2$ or $\sigma_1^2 < \cdots < \sigma_5^2$ or $\sigma_5^2 < \cdots < \sigma_1^2$) from the Math Garden example is specified as "not (1<2<3<4<5 or 5<4<3<2<1)" in the application (see Figure 9a).

Figure 9b shows the "Optional input" tab panel. Here the user may specify more advanced settings. Using the checkbox one can control whether the application shows Bayes factors or log Bayes factors (the latter is sometimes also referred to as the weight of evidence). In the next field the user may specify prior probabilities of the hypotheses. By default (i.e., if the field is empty) the posterior probabilities of the hypotheses are computed assuming equal prior probabilities. The "Fractions" field can be used to specify custom fractions $b_1, \ldots, b_J$. If the field is empty, the application uses the minimal information approach and sets $b_j =$

(a) The garden page (b) An addition item

Figure 8. Two screenshots of the Math Garden. Figure (a) shows the garden page where each plant represents a game measuring a different aspect of mathematics. Figure (b) shows an exemplary addition item. See the online article for the color version of this figure.

$2/n_j$ by default. Computing the marginal likelihood under an inequality constrained hypothesis involves sampling from the posterior and the prior distribution of the group variances. In the field "Number of simulation draws" one may specify how often to draw from the posterior and the prior. By default the application simulates 100,000 draws. In the last field the user may specify a custom seed in order to reproduce results exactly in the case of testing inequality constrained hypotheses (which requires simulating from the posterior and the prior). The "Help" tab panel contains detailed instructions on how to use the application.

Once all input has been specified, clicking on the "Submit" button initiates the computation of the results. Computation time mostly depends on the number of simulation draws, the number of hypotheses, and the number of inequality constraints. For example, the analysis shown in the screenshots should be completed within a few seconds. The results are shown in the output on the right-hand side of Figure 9a. The output consists of two tables, one showing the (log) Bayes factors and one showing the posterior probabilities of the hypotheses. The screenshot shows the results for the addition domain in the Math Garden example. In the "Bayes factors" table, the cell in row $t \in \{1, 2, 3, 4\}$ and column $t' \in \{1, 2, 3, 4\}$ contains the logarithm of the Bayes factor $B^{aF}_{tt'}$ (because we ticked the "Show logarithm of Bayes factors" checkbox in the optional input, see Figure 9b). For example, the cell in row 4 and column 1 contains the logarithm of the Bayes factor B^{aF}_{41} testing H_4 : not $(\sigma^2_1 < \cdots < \sigma^2_5$ or $\sigma^2_5 < \cdots < \sigma^2_1)$ against $H_1 : \sigma^2_1 = \cdots = \sigma^2_5$ (note that the hypotheses are numbered consecutively starting with 1). The log Bayes factor equals 251.33, which means that the evidence in the data in favor of H_4 is exp(251.33) times as strong as the evidence in favor of H_1. Some log Bayes factors are infinite because the marginal likelihoods under $H_2 : \sigma^2_1 < \cdots < \sigma^2_5$ and $H_3 : \sigma^2_5 < \cdots < \sigma^2_1$ are approximated as 0. As a result, the logarithms of the Bayes factors B^{aF}_{23} and B^{aF}_{32} are undefined, which is why in the corresponding cells in the table it says NA (for "not available").

To run our Shiny application follow these six steps:

1. Download and install R from https://cran.r-project.org/.

2. Launch R.

3. Copy the following R code and paste it into the R console:
```
install.packages("shiny")
```
Hit the Enter key and select a mirror. This will install the Shiny package.

4. Copy the following R code and paste it into the R console:
```
library(shiny)
```
Hit the Enter key. This will load the Shiny package.

5. Copy the following R code and paste it into the R console:
```
runGitHub("BFtestvar", "fboeingmessing")
```


Table 5
Descriptive Statistics for the Math Garden Data

	Sample size				Sample variance				Variance ratio			
Grade	+	−	×	÷	+	−	×	÷	+	−	×	÷
1	4,336	1,471	—	—	7.22	7.45	—	—	1.25	1.17	—	—
2	4,080	2,663	—	—	5.76	6.35	—	—	1.00	1.00	—	—
3	2,396	1,763	3,567	1,434	7.26	9.76	4.69	24.20	1.26	1.54	1.00	1.00
4	1,551	1,123	2,968	1,907	9.86	13.83	8.04	27.10	1.71	2.18	1.71	1.12
5	1,239	756	2,197	1,815	14.57	16.69	12.99	33.99	2.53	2.63	2.77	1.40
6	—	—	1,094	1,117	—	—	20.64	45.65	—	—	4.40	1.89

Note. The symbols +, −, ×, and ÷ refer to the corresponding domain in the Math Garden. The variance ratio is the ratio of the sample variance to the smallest sample variance in the corresponding domain.

Table 6
Results of the Analysis of the Math Garden Data

Result	$+$	$-$	$\times$	$\div$
$P^{aF}\,(H_0 : \sigma_1^2 = \cdots = \sigma_J^2 \,\vert\, \mathbf{x}, \mathbf{b})$	.00	.00	.00	.00
$P^{aF}\,(H_1 : \sigma_1^2 < \cdots < \sigma_J^2 \,\vert\, \mathbf{x}, \mathbf{b})$	.00	.03	1.00	1.00
$P^{aF}\,(H_2 : \sigma_J^2 < \cdots < \sigma_1^2 \,\vert\, \mathbf{x}, \mathbf{b})$	.00	.00	.00	.00
$P^{aF}\,(H_3 : \text{not } (H_0 \text{ or } H_1 \text{ or } H_2) \,\vert\, \mathbf{x}, \mathbf{b})$	1.00	.97	.00	.00

Note. The symbols $+$, $-$, $\times$, and $\div$ refer to the corresponding domain in the Math Garden.

Hit the Enter key. The Shiny application will open in your browser.

6. When you have completed your analyses you need to stop the application in order to be able to close R. To do so click on the red “STOP” button in the R menu bar.

Note that Steps 1 and 3 only need to be performed the first time you use the application. The R source code of the application is available at https://github.com/fboeingmessing/BFtestvar.

Discussion

In this article we developed a Bayes factor for testing equality and inequality constrained hypotheses on variances. Our method is based on an adjustment of the fractional Bayes factor (O’Hagan, 1995) such that it properly incorporates the parsimony of inequality constrained hypotheses. Using our adjusted fractional Bayes factor we can test any combination of equality and inequality constraints on the variances. It is straightforward to simultaneously test multiple hypotheses. The aFBF then indicates which hypothesis receives strongest support from the data. In doing so it functions as Occam’s razor by taking the parsimony of (in)equality constrained hypotheses into account. The aFBF is fully automatic, which means that the user does not need to specify a prior distribution under every hypothesis to be tested. The results of the simulation study indicate that the aFBF is consistent in the sense that it selects the true hypothesis as the sample size increases. This also holds for instances in which the true order of the population variances is slightly different from the hypothesized order. In this case the aFBF chooses the complement over the order constrained hypothesis as the sample size increases. The aFBF can be computed easily and quickly using our Shiny application.

In the multiplication and the division domain of the Math Garden the variances increased monotonically across grades as suggested by Aunola et al. (2004). In the addition and the subtraction domain, however, the variances first decreased from Grade 1 to Grade 2, followed by an increase over the years. Interestingly, both patterns are in line with a random slope model of development over time. Our approach can be used to test these and other variance patterns implied by models of development over time such as random slope and random quadratic models using cross-sectional data.

Like many other statistical methods, the aFBF assumes that the data are normally distributed. However, the normal distribution

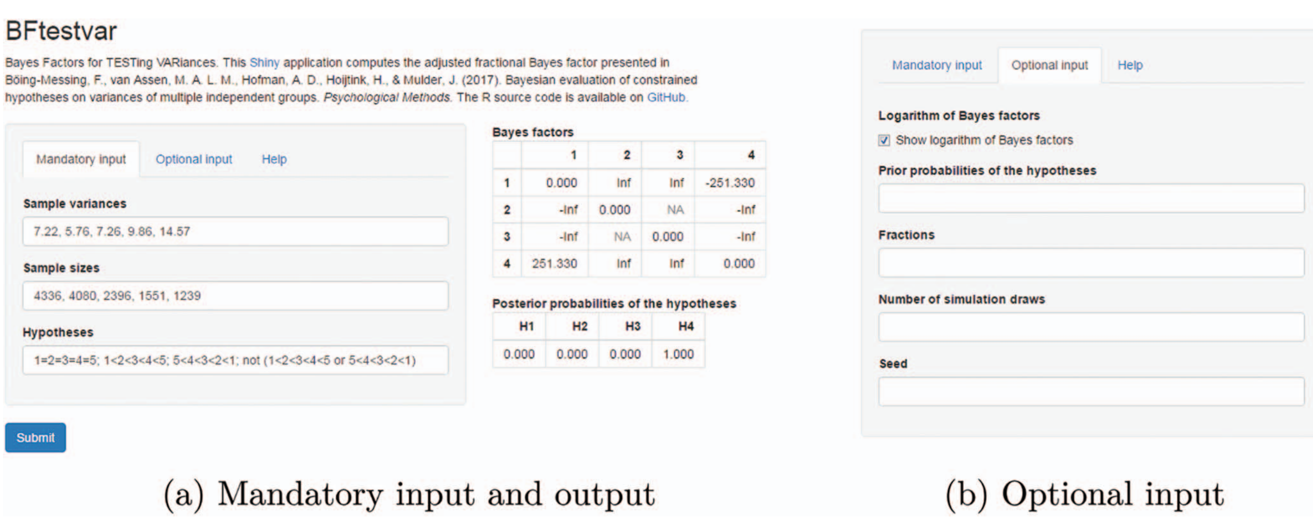

Figure 9. Two screenshots of the Shiny application for computing the adjusted fractional Bayes factor. Figure (a) shows the “Mandatory input” tab panel and the output (Bayes factors and posterior probabilities of the hypotheses). Figure (b) shows the “Optional input” tab panel. See the online article for the color version of this figure.

may not be an appropriate model for data that contain outliers or depart in other ways from normality (e.g., skewness and/or kurtosis). The robustness of the aFBF to such violations of normality is an important topic for future study. Furthermore, it would be interesting to investigate how the aFBF behaves under conditions that differ from those in our simulation study. For example, in the simulation we assumed that the ratio of adjacent population variances is constant. Real-life psychological phenomena may involve more complex variance patterns, which is why further research investigating the behavior of the aFBF under different population variance structures is indicated.

In this article we focused on testing variances of independent groups. It appears natural to also consider the Bayes factor for testing variances of dependent groups because these are frequently encountered by psychologists. Such a method would be useful for analyzing repeated measurement data and other types of data where there is a relationship between the respondents of different groups. Our approach can be extended to dependent observations using a multivariate normal model $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix of the dependent measures. Constrained hypotheses are then formulated on the diagonal elements of this covariance matrix. The additional challenge in the dependent case is that the constraints on the variances are added to the constraints that ensure that the covariance matrix is positive definite. This is an interesting topic for future research.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *2nd International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akadémiai Kiadó.

Anraku, K. (1999). An information criterion for parameters under a simple order restriction. *Biometrika, 86,* 141–152.

Aunola, K., Leskinen, E., Lerkkanen, M.-K., & Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96,* 699–713.

Bartlett, M. S. (1957). A comment on D. V. Lindley's statistical paradox. *Biometrika, 44,* 533–534.

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis, 1,* 385–402.

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

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

Böing-Messing, F., & Mulder, J. (2016). Automatic Bayes factors for testing and. *Journal of Mathematical Psychology, 72,* 158–170.

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,* 49–60.

Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin, 104,* 396–404.

Carroll, R. J. (2003). Variances are not always nuisance parameters. *Biometrics, 59,* 211–220.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2015). *shiny: Web application framework for R*. Retrieved from http://CRAN.R-project.org/package=shiny

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

De Santis, F., & Spezzaferri, F. (2001). Consistent fractional Bayes factor for nested normal linear models. *Journal of Statistical Planning and Inference, 97,* 305–321.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160.

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

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science, 24,* 343–360.

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

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24,* 997–1016.

Gilks, W. R. (1995). Discussion of O'Hagan. *Journal of the Royal Statistical Society Series B, 57,* 118–120.

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of Consulting and Clinical Psychology, 68,* 155–165.

Gu, X., Mulder, J., Deković, M., & Hoijtink, H. (2014). Bayesian evaluation of inequality constrained hypotheses. *Psychological Methods, 19,* 511–527.

Hoijtink, H. (2011). *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: Chapman & Hall/CRC.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journal of Gerontology: Psychological Sciences, 57B,* 101–115.

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.

Kim, J., & Seltzer, M. (2011). Examining heterogeneity in residual variance to detect differential response to treatments. *Psychological Methods, 16,* 192–208.

Klinkenberg, S., Straatemeier, M., & van der Maas, H. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education, 57,* 1813–1824.

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

Lehre, A.-C., Lehre, K. P., Laake, P., & Danbolt, N. C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology, 51,* 198–206.

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, & H. B. Mann (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford, CA: Stanford University Press.

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

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

Maris, G., & van der Maas, H. (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika, 77,* 615–633.

Mulder, J. (2014a). Bayes factors for testing inequality constrained hypotheses: Issues with prior specification. *British Journal of Mathematical and Statistical Psychology, 67,* 153–171.
Mulder, J. (2014b). 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,* 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,* 887–906.
Mulder, J., Klugkist, I., Schoot, R., van de Meeus, W. H., Selfhout, M., & Hoijtink, H. (2009). Bayesian model selection of informative hypotheses for repeated measurements. *Journal of Mathematical Psychology, 53,* 530–546.
O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society Series B, 57,* 99–138.
O'Hagan, A. (1997). Properties of intrinsic and fractional Bayes factors. *Test, 6,* 101–118.
Pérez, J. M., & Berger, J. O. (2002). Expected-posterior prior distributions for model selection. *Biometrika, 89,* 491–511.
R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
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,* 225–237.
Ruscio, J., & Roche, B. (2012). Variance heterogeneity in published psychological research: A review and a new index. *Methodology, 8,* 1–11.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464.
Silverstein, S. M., Como, P. G., Palumbo, D. R., West, L. L., & Osborn, L. M. (1995). Multiple sources of attentional dysfunction in adults with Tourette's syndrome: Comparison with attention deficit-hyperactivity disorder. *Neuropsychology, 9,* 157–164.
Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11,* 54–71.
Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London, UK: Sage.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, 64,* 583–639.
Straatemeier, M. (2014). *Math Garden: A new educational and scientific instrument* (Doctoral dissertation). University of Amsterdam, Amsterdam, the Netherlands. Retrieved from http://dare.uva.nl/record/1/417091
Van Mechelen, I. (2009). A royal road to understanding the mechanisms underlying person-in-context behavior. *Journal of Research in Personality, 43,* 179–186.
Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathematical and Statistical Psychology, 66,* 383–401.
Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804.
Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11,* 3571–3594.

Appendix A

Fractional Bayes Factor for an Inequality Constrained Hypothesis Test

We consider the test of an inequality constrained hypothesis H_t on the variances of J groups against the unconstrained hypothesis $H_u : \sigma_1^2, \ldots, \sigma_J^2$. The inequality constrained hypothesis can be formulated as $H_t : \mathbf{R}_t\boldsymbol{\sigma}^2 > \mathbf{0}$, where the rows of $\mathbf{R}_t$ are permutations of $(1, -1, 0, \ldots, 0)$. For example, under $H_1 : \sigma_1^2 < \sigma_2^2 < \sigma_3^2$ the matrix is given by $\mathbf{R}_1 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$. The admissible parameter space of the group variances under H_t and H_u can be written as $\Omega_t = \{\boldsymbol{\sigma}^2 \mid \mathbf{R}_t\boldsymbol{\sigma}^2 > \mathbf{0}\}$ and $\Omega_u = (\mathbb{R}^+)^J$, respectively. Note that the likelihood and the noninformative improper prior under H_t are truncations of the unconstrained likelihood and prior:

$$f_t(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = f_u(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2) I_{\Omega_t}(\boldsymbol{\sigma}^2), \quad (A.1)$$

$$f_t(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b = f_u(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b I_{\Omega_t}(\boldsymbol{\sigma}^2), \text{ and} \quad (A.2)$$

$$\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) = C_t \pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2) I_{\Omega_t}(\boldsymbol{\sigma}^2), \quad (A.3)$$

where $I_{\Omega_t}(\boldsymbol{\sigma}^2)$ is an indicator function that equals 1 if $\boldsymbol{\sigma}^2 \in \Omega_t$ and 0 otherwise, and C_t is a normalizing constant. The FBF for an inequality constrained hypothesis H_t against the unconstrained hypothesis H_u can then be written as

(*Appendices continue*)

$$B_{tu}^F = \frac{m_t^F(\mathbf{x}, b)}{m_u^F(\mathbf{x}, b)} = \frac{\dfrac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J} f_t(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_t^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}}{\dfrac{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}} = \frac{\dfrac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)C_t\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)I_{\Omega_t}(\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}}{\dfrac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^bC_t\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)I_{\Omega_t}(\boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}}$$

$$= \frac{\int_{\Omega_t}\int_{\mathbb{R}^J} \dfrac{f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J} \dfrac{f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)}{\int_{\Omega_u}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \boldsymbol{\sigma}^2)^b\pi_u^N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} = \frac{\int_{\Omega_t}\int_{\mathbb{R}^J} \pi_u(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\,|\,\mathbf{x})d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2}{\int_{\Omega_t}\int_{\mathbb{R}^J} \pi_u(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\,|\,\mathbf{x}^b)d\boldsymbol{\mu}\, d\boldsymbol{\sigma}^2} \tag{A.4}$$

$$= \frac{\int_{\Omega_t} \pi_u(\boldsymbol{\sigma}^2\,|\,\mathbf{x})d\boldsymbol{\sigma}^2}{\int_{\Omega_t} \pi_u(\boldsymbol{\sigma}^2\,|\,\mathbf{x}^b)d\boldsymbol{\sigma}^2} = \frac{P(\boldsymbol{\sigma}^2 \in \Omega_t\,|\,\mathbf{x})}{P(\boldsymbol{\sigma}^2 \in \Omega_t\,|\,\mathbf{x}^b)}.$$

Note that in the second line the indicator function $I_{\Omega_t}(\boldsymbol{\sigma}^2)$ in the constrained likelihood and prior, f_t and π_t^N, respectively, can be omitted because the integration region is already restricted to the constrained parameter space Ω_t.

Appendix B

Computation of the Marginal Likelihood in the Adjusted Fractional Bayes Factor

We consider a hypothesis H_t with equality and inequality constraints on the variances. We have to introduce some additional notation before deriving the marginal likelihood in the aFBF approach. Under H_t, let there be q^E equality constraints and q^I inequality constraints on the variances, where we omitted the hypothesis index t on q^E and q^I to simplify the notation. Thus, there are $K = J - q^E$ unique variances under H_t. We denote these K unique variances by $\tilde{\boldsymbol{\sigma}}^2 = (\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_K^2)'$. The q^I inequality constraints are formulated on these unique variances. Furthermore let J_k be the number of groups that share the unique variance $\tilde{\sigma}_k^2$, and let $\mathbf{x}_{k_j}$, μ_{k_j}, and n_{k_j} denote the data, the mean, and the sample size of the jth group sharing the unique variance $\tilde{\sigma}_k^2$, respectively.

For example, consider the hypothesis $H_1 : \sigma_1^2 = \sigma_2^2 < \sigma_3^2 = \sigma_4^2$ on the variances of $J = 4$ groups. Under H_1 there are $q^E = 2$ equality constraints and $q^I = 1$ inequality constraint, so that the number of unique variances is given by $K = 4 - 2 = 2$. We denote these variances by $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$. Then Groups 1 and 2 have unique variance $\tilde{\sigma}_1^2$ and Groups 3 and 4 have unique variance $\tilde{\sigma}_2^2$. Thus, hypothesis H_1 can be written as $H_1 : \tilde{\sigma}_1^2 < \tilde{\sigma}_2^2$. Furthermore, we have $(J_1, J_2)' = (2, 2)'$. In this notation, $\mathbf{x}_{1_1}$, $\mathbf{x}_{1_2}$, $\mathbf{x}_{2_1}$, and $\mathbf{x}_{2_2}$ correspond to the data of Groups 1, 2, 3, and 4, respectively, μ_{1_1}, μ_{1_2}, μ_{2_1}, and μ_{2_2} are the means of Groups 1, 2, 3, and 4, and n_{1_1}, n_{1_2}, n_{2_1}, and n_{2_2} are the sample sizes of Groups 1, 2, 3, and 4.

The marginal likelihood under a constrained hypothesis H_t in the adjusted fractional Bayes factor is defined by

$$m_t^{aF}(\mathbf{x}, \mathbf{b}) = \frac{\int_{\Omega_t}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2)\pi_u^N(\boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2)d\boldsymbol{\mu}\, d\tilde{\boldsymbol{\sigma}}^2}{\int_{\Omega_t^a}\int_{\mathbb{R}^J} f_u(\mathbf{x}\,|\,\boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2)^{\mathbf{b}}\pi_u^N(\boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2)d\boldsymbol{\mu}\, d\tilde{\boldsymbol{\sigma}}^2} = \frac{m_t^N(\mathbf{x})}{m_t^N(\mathbf{x}^{\mathbf{b}})}, \tag{B.1}$$

where the likelihood and the noninformative prior are used without the inequality constraints on the unique variances $\tilde{\boldsymbol{\sigma}}^2$, which is part of the definition of the aFBF. The expressions are given by

(*Appendices continue*)

$$f_u(\mathbf{x} \mid \boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2) = \prod_{k=1}^{K} \prod_{j=1}^{J_k} f\left(\mathbf{x}_{k_j} \mid \mu_{k_j}, \tilde{\sigma}_k^2\right), \tag{B.2}$$

$$f_u(\mathbf{x} \mid \boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2)^{\mathbf{b}} = \prod_{k=1}^{K} \prod_{j=1}^{J_k} f\left(\mathbf{x}_{k_j} \mid \mu_{k_j}, \tilde{\sigma}_k^2\right)^{b_{k_j}}, \text{ and} \tag{B.3}$$

$$\pi_u^N(\boldsymbol{\mu}, \tilde{\boldsymbol{\sigma}}^2) = C_t \prod_{k=1}^{K} \tilde{\sigma}_k^{-2}, \tag{B.4}$$

with

$$f\left(\mathbf{x}_{k_j} \mid \mu_{k_j}, \tilde{\sigma}_k^2\right) = \left(\tilde{\sigma}_k^2 2\pi\right)^{-\frac{n_{k_j}}{2}} \exp\left(-\frac{1}{2\tilde{\sigma}_k^2}\left((n_{k_j}-1)s_{k_j}^2 + n_{k_j}(\bar{x}_{k_j} - \mu_{k_j})^2\right)\right) \text{ and} \tag{B.5}$$

$$f\left(\mathbf{x}_{k_j} \mid \mu_{k_j}, \tilde{\sigma}_k^2\right)^{b_{k_j}} = \left(\tilde{\sigma}_k^2 2\pi\right)^{-\frac{b_{k_j} n_{k_j}}{2}} \exp\left(-\frac{b_{k_j}}{2\tilde{\sigma}_k^2}\left((n_{k_j}-1)s_{k_j}^2 + n_{k_j}(\bar{x}_{k_j} - \mu_{k_j})^2\right)\right). \tag{B.6}$$

Note that Equation (B.1) is identical to Equation (15) except that a tilde is used for the unique variances which are integrated out.

The constrained parameter space Ω_t in the numerator in Equation (B.1) can be written as

$$\Omega_t = \{\tilde{\boldsymbol{\sigma}}^2 \mid \mathbf{R}_t(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_K^2)' > \mathbf{0}\}, \tag{B.7}$$

where the rows of $\mathbf{R}_t$ are permutations of $(1, -1, 0, \ldots, 0)$. For example, under $H_1 : \tilde{\sigma}_1^2 < \tilde{\sigma}_2^2$ the matrix is given by $\mathbf{R}_1 = [-1 \quad 1]$. The adjusted constrained parameter space Ω_t^a in the denominator in Equation (B.1), which is a crucial part of the aFBF approach, can be written as

$$\Omega_t^a = \{\tilde{\boldsymbol{\sigma}}^2 \mid \mathbf{R}_t(a_1\tilde{\sigma}_1^2, \ldots, a_K\tilde{\sigma}_K^2)' > \mathbf{0}\}, \tag{B.8}$$

where the tuning parameters are set to

$$a_k = \frac{\left(\sum_{j=1}^{J_k} b_{k_j} n_{k_j}\right) - J_k}{\sum_{j=1}^{J_k} b_{k_j}\left(n_{k_j} - 1\right)s_{k_j}^2}, \tag{B.9}$$

for $k = 1, \ldots, K$. This tuning results in equal scale hyperparameters in the automatic prior for the unique variances. This will be shown after the derivation of the marginal likelihood.

We first derive the denominator $m_t^N(\mathbf{x}^{\mathbf{b}})$ of the marginal likelihood in Equation (B.1). Substituting the expressions for the fraction of the likelihood and the Jeffreys prior in Equations (B.3) and (B.4) into the denominator of Equation (B.1) gives us

(Appendices continue)

$$
\begin{aligned}
m_t^N(\mathbf{x^b}) &= \int_{\Omega_t^a}\int_{\mathbb{R}^J}\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} f\left(\mathbf{x}_{k_j}\,|\,\mu_{k_j},\tilde{\sigma}_k^2\right)^{b_{k_j}}\right)C_t\prod_{k=1}^{K}\tilde{\sigma}_k^{-2}d\boldsymbol{\mu}d\tilde{\boldsymbol{\sigma}}^2\\
&= C_t\int_{\Omega_t^a}\prod_{k=1}^{K}\tilde{\sigma}_k^{-2}\prod_{j=1}^{J_k}\int_{\mathbb{R}}(\tilde{\sigma}_k^2 2\pi)^{-\frac{b_{k_j}n_{k_j}}{2}}\\
&\quad \exp\left(-\frac{b_{k_j}}{2\tilde{\sigma}_k^2}\left((n_{k_j}-1)s_{k_j}^2+n_{k_j}(\bar{x}_{k_j}-\mu_{k_j})^2\right)\right)d\mu_{k_j}d\tilde{\boldsymbol{\sigma}}^2\\
&= C_t\int_{\Omega_t^a}\prod_{k=1}^{K}\tilde{\sigma}_k^{-2}\prod_{j=1}^{J_k}(b_{k_j}n_{k_j})^{-\frac{1}{2}}(\tilde{\sigma}_k^2 2\pi)^{-\frac{b_{k_j}n_{k_j}-1}{2}}\exp\left(-\frac{b_{k_j}(n_{k_j}-1)s_{k_j}^2}{2\tilde{\sigma}_k^2}\right)d\tilde{\boldsymbol{\sigma}}^2\\
&= C_t\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k}(b_{k_j}n_{k_j})^{-\frac{1}{2}}\right)(2\pi)^{-\frac{\sum_{k=1}^{K}\left(\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k\right)}{2}}\\
&\quad \int_{\Omega_t^a}\prod_{k=1}^{K}(\tilde{\sigma}_k^2)^{-\left(\frac{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}{2}+1\right)}\exp\left(-\frac{\sum_{j=1}^{J_k}b_{k_j}(n_{k_j}-1)s_{k_j}^2}{2\tilde{\sigma}_k^2}\right)d\tilde{\boldsymbol{\sigma}}^2\\
&= C_t\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k}(b_{k_j}n_{k_j})^{-\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\left(\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k\right)}{2}}\\
&\quad \left(\prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}{2}\right)\left(\sum_{j=1}^{J_k}b_{k_j}(n_{k_j}-1)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}{2}}\right)\\
&\quad\quad \int_{\Omega_t^a}\prod_{k=1}^{K}\text{Inv-}\chi^2\left(\tilde{\sigma}_k^2\,\middle|\,\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k,\frac{\sum_{j=1}^{J_k}b_{k_j}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}\right)d\tilde{\boldsymbol{\sigma}}^2\\
&= C_t\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k}(b_{k_j}n_{k_j})^{-\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\left(\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k\right)}{2}}\\
&\quad \left(\prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}{2}\right)\left(\sum_{j=1}^{J_k}b_{k_j}(n_{k_j}-1)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}{2}}\right)\\
&\quad P(\tilde{\boldsymbol{\sigma}}^2\in\Omega_t^a\,|\,\mathbf{x^b}).
\end{aligned} \tag{B.10}
$$

In the second line we solved the integral with respect to μ_{k_j} by integrating $\exp\left(-\frac{b_{k_j}n_{k_j}}{2\tilde{\sigma}_k^2}(\mu_{k_j}-\bar{x}_{k_j})^2\right)$, which is the kernel of a normal distribution with mean $\bar{x}_{k_j}$ and variance $\tilde{\sigma}_k^2/(b_{k_j}n_{k_j})$. Hence the integral equals $(2\pi\tilde{\sigma}_k^2/(b_{k_j}n_{k_j}))^{\frac{1}{2}}$, which is the inverse of the normalizing constant of this normal distribution. The integrand in the fourth line is a product of kernels of scaled inverse-χ^2 distributions with degrees of freedom parameters $\nu_k = (\sum_{j=1}^{J_k}b_{k_j}n_{k_j}) - J_k$ and scale parameters $\tau_k^2 = \frac{\sum_{j=1}^{J_k}b_{k_j}(n_{k_j}-1)s_{k_j}^2}{(\sum_{j=1}^{J_k}b_{k_j}n_{k_j})-J_k}$, $k = 1, \ldots, K$ (Gelman et al., 2004). Finally, the probability that the variances fall in the adjusted parameter space Ω_t^a is based on independent automatic priors for the variances given by

$$
\tilde{\sigma}_k^2\,|\,\mathbf{x^b}\sim\ \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k,\frac{\sum_{j=1}^{J_k}b_{k_j}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k}b_{k_j}n_{k_j}\right)-J_k}\right), \tag{B.11}
$$

for $k = 1, \ldots, K$.

(*Appendices continue*)

The expression for the numerator $m_t^N(\mathbf{x})$ of the marginal likelihood in Equation (B.1) is identical to the final expression in Equation (B.10) with all b's set to 1 and Ω_t^a replaced by Ω_t:

$$\begin{aligned} m_t^N(\mathbf{x}) &= C_t\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} n_{k_j}^{-\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k\right)}{2}} \\ &\left(\prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\left(\sum_{j=1}^{J_k}(n_{k_j}-1)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}}\right) \\ &P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid \mathbf{x}). \end{aligned} \tag{B.12}$$

Subsequently, the marginal likelihood in the aFBF is given by

$$m_t^{aF}(\mathbf{x},\mathbf{b}) = \frac{m_t^N(\mathbf{x})}{m_t^N(\mathbf{x}^{\mathbf{b}})} = \widetilde{m}_t^{aF}(\mathbf{x},\mathbf{b})\frac{P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid \mathbf{x})}{P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}})}, \tag{B.13}$$

where

$$\begin{aligned} \widetilde{m}_t^{aF}(\mathbf{x},\mathbf{b}) = &\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}(1-b_{k_j})n_{k_j}}{2}} \\ &\prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}\right)^{-1} \\ &\left(\sum_{j=1}^{J_k}(n_{k_j}-1)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}}\left(\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}}. \end{aligned} \tag{B.14}$$

Note that if a constrained hypothesis does not contain any inequalities, the ratio of probabilities in Equation (B.13) is not present.

Finally, we provide a motivation for the specific choice of the tuning parameters. First we introduce new parameters $\phi_k = a_k\tilde{\sigma}_k^2$, for $k = 1, \ldots, K$, which can be interpreted as adjusted variance parameters. In the automatic prior the adjusted variance is distributed according to

$$\begin{aligned} \phi_k = a_k\tilde{\sigma}_k^2 \mid \mathbf{x}^{\mathbf{b}} &\sim \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k,\, a_k\frac{\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}\right) \\ &= \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k,\, 1\right), \end{aligned} \tag{B.15}$$

for $k = 1, \ldots, K$, which follows automatically from Equations (B.9) and (B.11). In the first line we used the mathematical result that if $\tilde{\sigma}_k^2 \mid \mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2(\nu_k, \tau_k^2)$, then $a_k\tilde{\sigma}_k^2 \mid \mathbf{x}^{\mathbf{b}} \sim \text{Inv-}\chi^2(\nu_k, a_k\tau_k^2)$. Note that the scale hyperparameters of the scaled inverse-χ^2 distributions are equal for all k. Subsequently, the automatic prior probability that the variances fall in the adjusted constrained space Ω_t^a can be written as

$$P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}}) = P(\mathbf{R}_t(a_1\tilde{\sigma}_1^2, \ldots, a_K\tilde{\sigma}_K^2)' > \mathbf{0} \mid \mathbf{x}^{\mathbf{b}}) = P(\mathbf{R}_t(\phi_1, \ldots, \phi_K)' > \mathbf{0} \mid \mathbf{x}^{\mathbf{b}}). \tag{B.16}$$

To illustrate the effect of the adjustment we again consider the hypothesis $H_1: \tilde{\sigma}_1^2 < \tilde{\sigma}_2^2$ with $\mathbf{R}_1 = [-1 \quad 1]$. If we set $b_{k_j} = 2/n_{k_j}$, the automatic prior probability that the variances fall in the adjusted constrained space equals

$$P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}}) = P(a_1\tilde{\sigma}_1^2 < a_2\tilde{\sigma}_2^2 \mid \mathbf{x}^{\mathbf{b}}) = P(\phi_1 < \phi_2 \mid \mathbf{x}^{\mathbf{b}}) = \frac{1}{2}, \tag{B.17}$$

because ϕ_1 and ϕ_2 are both distributed as Inv-$\chi^2(2, 1)$ due to Equation (B.15). This is desirable because it implies that in the aFBF approach both possible orderings of the two adjusted variances are equally likely a priori.

(*Appendices continue*)

Appendix C

Scale Invariance of the Adjusted Fractional Bayes Factor

In Appendix B the data and the sample variance of the jth group sharing the unique variance $\tilde{\sigma}_k^2$ were denoted by $\mathbf{x}_{k_j}$ and $s_{k_j}^2$, respectively. Multiplying all observations in $\mathbf{x}_{k_j}$ by a constant w results in a sample variance of $w^2 s_{k_j}^2$, for $j = 1, \ldots, J_k$ and $k = 1, \ldots, K$. We show that the marginal likelihood of the scaled data $w\mathbf{x}$ under hypothesis H_t can be written as $m_t^{aF}(w\mathbf{x}, \mathbf{b}) = v m_t^{aF}(\mathbf{x}, \mathbf{b})$, where v is a constant that is independent of H_t. As will be shown, v cancels out in the computation of the adjusted fractional Bayes factors and the corresponding posterior probabilities of the hypotheses. We first consider $\widetilde{m}_t^{aF}(w\mathbf{x}, \mathbf{b})$ in Equation (B.13) for the scaled data. Note that the marginal likelihood only depends on the data through the sample variances. Thus, substituting $s_{k_j}^2$ with $w^2 s_{k_j}^2$ in Equation (B.14) gives us

$$\begin{aligned}
\widetilde{m}_t^{aF}(w\mathbf{x}, \mathbf{b}) &= \left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}(1-b_{k_j})n_{k_j}}{2}} \\
&\quad \prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}\right)^{-1} \\
&\quad \left(\sum_{j=1}^{J_k}(n_{k_j}-1)w^2 s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}}\left(\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)w^2 s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}} \\
&= (w^2)^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}(1-b_{k_j})n_{k_j}}{2}}\left(\prod_{k=1}^{K}\prod_{j=1}^{J_k} b_{k_j}^{\frac{1}{2}}\right)\pi^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}(1-b_{k_j})n_{k_j}}{2}} \\
&\quad \prod_{k=1}^{K}\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}\right)\Gamma\left(\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}\right)^{-1} \\
&\quad \left(\sum_{j=1}^{J_k}(n_{k_j}-1)s_{k_j}^2\right)^{-\frac{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}{2}}\left(\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)s_{k_j}^2\right)^{\frac{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}{2}} \\
&= (w^2)^{-\frac{\sum_{k=1}^{K}\sum_{j=1}^{J_k}(1-b_{k_j})n_{k_j}}{2}}\widetilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) = v\widetilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}).
\end{aligned} \tag{C.1}$$

Next, we consider $P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid w\mathbf{x})$ and $P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid (w\mathbf{x})^{\mathbf{b}})$ in Equation (B.13) for the scaled data. For the scaled data the variances are distributed according to

$$\tilde{\sigma}_k^2 \mid (w\mathbf{x})^{\mathbf{b}} \sim \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right) - J_k, w^2\frac{\sum_{j=1}^{J_k} b_{k_j}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} b_{k_j}n_{k_j}\right)-J_k}\right) \quad \text{and} \tag{C.2}$$

$$\tilde{\sigma}_k^2 \mid w\mathbf{x} \sim \text{Inv-}\chi^2\left(\left(\sum_{j=1}^{J_k} n_{k_j}\right) - J_k, w^2\frac{\sum_{j=1}^{J_k}(n_{k_j}-1)s_{k_j}^2}{\left(\sum_{j=1}^{J_k} n_{k_j}\right)-J_k}\right), \tag{C.3}$$

for the automatic prior and posterior, respectively, for $k = 1, \ldots, K$. Because the scale parameters in the above distributions only depend on the scale w through the factor w^2, it automatically follows that the automatic prior probability is invariant of the scale, that is,

$$\begin{aligned}
P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid (w\mathbf{x})^{\mathbf{b}}\right) &= P\left(w^2\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}}\right) = P\left(\mathbf{R}_t(w^2 a_1\tilde{\sigma}_1^2, \ldots, w^2 a_K\tilde{\sigma}_K^2)' > \mathbf{0} \mid \mathbf{x}^{\mathbf{b}}\right) \\
&= P\left(\mathbf{R}_t(a_1\tilde{\sigma}_1^2, \ldots, a_K\tilde{\sigma}_K^2)' > \mathbf{0} \mid \mathbf{x}^{\mathbf{b}}\right) = P\left(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}}\right).
\end{aligned} \tag{C.4}$$

(*Appendices continue*)

By following the same steps it can be shown that the posterior probability is also invariant, that is, $P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid w\mathbf{x}) = P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid \mathbf{x})$. Thus, the marginal likelihood of the scaled data can be written as

$$\begin{aligned} m_t^{aF}(w\mathbf{x}, \mathbf{b}) &= \widetilde{m}_t^{aF}(w\mathbf{x}, \mathbf{b}) \frac{P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid w\mathbf{x})}{P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid (w\mathbf{x})^{\mathbf{b}})} \\ &= v\widetilde{m}_t^{aF}(\mathbf{x}, \mathbf{b}) \frac{P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t \mid \mathbf{x})}{P(\tilde{\boldsymbol{\sigma}}^2 \in \Omega_t^a \mid \mathbf{x}^{\mathbf{b}})} = v m_t^{aF}(\mathbf{x}, \mathbf{b}). \end{aligned} \tag{C.5}$$

Because the constant v is the same under all hypotheses, it cancels out in the computation of the adjusted fractional Bayes factors and the corresponding posterior probabilities of the hypotheses:

$$B_{tt'}^{aF} = \frac{m_t^{aF}(w\mathbf{x}, \mathbf{b})}{m_{t'}^{aF}(w\mathbf{x}, \mathbf{b})} = \frac{v m_t^{aF}(\mathbf{x}, \mathbf{b})}{v m_{t'}^{aF}(\mathbf{x}, \mathbf{b})} = \frac{m_t^{aF}(\mathbf{x}, \mathbf{b})}{m_{t'}^{aF}(\mathbf{x}, \mathbf{b})} \tag{C.6}$$

and

$$\begin{aligned} P^{aF}(H_t \mid w\mathbf{x}, \mathbf{b}) &= \frac{m_t^{aF}(w\mathbf{x}, \mathbf{b}) P(H_t)}{\sum_{t'=1}^{T} m_{t'}^{aF}(w\mathbf{x}, \mathbf{b}) P(H_{t'})} = \frac{v m_t^{aF}(\mathbf{x}, \mathbf{b}) P(H_t)}{\sum_{t'=1}^{T} v m_{t'}^{aF}(\mathbf{x}, \mathbf{b}) P(H_{t'})} \\ &= \frac{m_t^{aF}(\mathbf{x}, \mathbf{b}) P(H_t)}{\sum_{t'=1}^{T} m_{t'}^{aF}(\mathbf{x}, \mathbf{b}) P(H_{t'})} = P^{aF}(H_t \mid \mathbf{x}, \mathbf{b}). \end{aligned} \tag{C.7}$$

Received December 24, 2015
Revision received August 31, 2016
Accepted September 9, 2016 ■