



Bayes Factors for Testing Order Constraints on Variances of Dependent Outcomes

Florian Böing-Messing & Joris Mulder

To cite this article: Florian Böing-Messing & Joris Mulder (2020): Bayes Factors for Testing Order Constraints on Variances of Dependent Outcomes, The American Statistician, DOI: [10.1080/00031305.2020.1715257](https://doi.org/10.1080/00031305.2020.1715257)

To link to this article: <https://doi.org/10.1080/00031305.2020.1715257>



© 2020 The Author(s). Published with license by Taylor and Francis Group, LLC



Published online: 12 Feb 2020.



Submit your article to this journal [↗](#)



Article views: 576



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Bayes Factors for Testing Order Constraints on Variances of Dependent Outcomes

Florian Böing-Messing^{a,b} and Joris Mulder^{b,a}

^aJheronimus Academy of Data Science, 's-Hertogenbosch, The Netherlands; ^bDepartment of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

ABSTRACT

In statistical practice, researchers commonly focus on patterns in the means of multiple dependent outcomes while treating variances as nuisance parameters. However, in fact, there are often substantive reasons to expect certain patterns in the variances of dependent outcomes as well. For example, in a repeated measures study, one may expect the variance of the outcome to increase over time if the difference between subjects becomes more pronounced over time because the subjects respond differently to a given treatment. Such expectations can be formulated as order constrained hypotheses on the variances of the dependent outcomes. Currently, however, no methods exist for testing such hypotheses in a direct manner. To fill this gap, we develop a Bayes factor for this challenging testing problem. Our Bayes factor is based on the multivariate normal distribution with an unstructured covariance matrix, which is often used to model dependent outcomes. Order constrained hypotheses can then be formulated on the variances on the diagonal of the covariance matrix. To compute Bayes factors between multiple order constrained hypotheses, a prior distribution needs to be specified under every hypothesis to be tested. Here, we use the encompassing prior approach in which priors under order constrained hypotheses are truncations of the prior under the unconstrained hypothesis. The resulting Bayes factor is fully automatic in the sense that no subjective priors need to be specified by the user.

ARTICLE HISTORY

Received May 2018
Accepted December 2019

KEYWORDS

Encompassing prior; Heteroscedasticity; Inequality constraint; Mixed model; Multivariate regression; Repeated measures

1. Introduction

In analyzing their data, researchers commonly focus on measures of central tendency such as means and regression coefficients. Measures such as variances that capture the heterogeneity of observations are usually regarded as nuisance parameters. However, to fully understand the patterns that are present in the data it is of vital importance to carefully model and interpret the variability of the observations (e.g., Carroll 2003). In fact, the heterogeneity of the observations can be a core aspect of a study (e.g., Hultsch, MacDonald, and Dixon 2002; Aunola et al. 2004; Lehre et al. 2009; Kaptein and Eckles 2012; Mulder and Fox 2013; Böing-Messing and Mulder 2016, 2018; Mulder and Fox 2019).

There are often reasons to expect certain patterns in the heterogeneity of multiple dependent outcomes. For example, in a repeated measures study that investigates the effect of a certain treatment, one may expect the variability of the outcome to increase over time if the difference between subjects becomes more pronounced over time because the subjects respond differently to the treatment (e.g., Aunola et al. 2004; Hedeker and Gibbons 2006; Böing-Messing et al. 2017). Such an expectation can be formulated as an order constrained hypothesis of the form $H_1: \sigma_1^2 < \dots < \sigma_p^2$ subject to the condition that the covariance matrix is positive definite, where p is the number of measurement occasions and σ_j^2 is the variance of the j th measurement, for $j = 1, \dots, p$. To test H_1 we need a hypothesis it can be compared to. A natural competitor is the complement

of H_1 given by $H_2: \neg \sigma_1^2 < \dots < \sigma_p^2$, which may also be written as $H_2: \neg H_1$ in short. The complement entails all possible hypotheses on the p variances except H_1 under the condition that the covariance matrix is positive definite. In another example, Aunola et al. (2004) hypothesized that the variability of math performances either increases or decreases across grades. An increase might occur because children with high mathematical potential improve their performances faster than children with low mathematical potential. A possible reason for a decrease is that systematic instruction at school helps children with low mathematical potential catch up. These two competing expectations can be expressed as order constrained hypotheses $H_1: \sigma_1^2 < \dots < \sigma_p^2$ and $H_2: \sigma_p^2 < \dots < \sigma_1^2$, where p is the number of grades and σ_j^2 is the variance in grade j , for $j = 1, \dots, p$. Another conceivable competitor is the complement of H_1 and H_2 given by $H_3: \neg (\sigma_1^2 < \dots < \sigma_p^2 \vee \sigma_p^2 < \dots < \sigma_1^2)$.

In this article, we use the multivariate normal distribution with unstructured $p \times p$ covariance matrix Σ to model the dependent outcomes. We use an unstructured covariance matrix to safeguard against misfitting covariance structures that might distort inferences about the variances. As we will argue later, an unstructured covariance matrix is more suitable for testing specific patterns of the variances in comparison to the implied, highly structured covariance matrix of a mixed effects model. We concern ourselves with testing $T \geq 2$ order constrained hypotheses on the variances $\sigma^2 = (\sigma_1^2, \dots, \sigma_p^2)'$ of the dependent outcomes

located on the main diagonal of Σ . In general, the hypotheses we test are of the form

$$H_t: \Sigma \in \Omega_t, \quad t = 1, \dots, T, \quad (1)$$

where Ω_t is an order constrained subspace of the unconstrained parameter space $\Omega_u = \{\Sigma : \Sigma > 0\}$ and $\Sigma > 0$ means that Σ is positive definite. In this article, we focus on hypotheses for which the subspace can be written as $\Omega_t = \{\Sigma : \mathbf{R}_t \sigma^2 > \mathbf{0} \wedge \Sigma > 0\}$, where \mathbf{R}_t is a $q_t \times p$ matrix containing the coefficients for the q_t order constraints on the variances under H_t and $\mathbf{0} = (0, \dots, 0)'$ is a q_t -dimensional vector of zeroes. We test hypotheses for which each row of \mathbf{R}_t is a permutation of $\{-1, 1, 0, \dots, 0\}$. That is, we consider hypotheses with equal coefficients on the variances (e.g., $H_1: \Sigma \in \Omega_1 = \{\Sigma : \sigma_1^2 < \sigma_2^2 < \sigma_3^2 \wedge \Sigma > 0\}$, for which $\mathbf{R}_1 = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$). Another type of hypothesis we will consider in the context of repeated measures analysis is a formulation that states an order of the variances as well as an order of the ratios of adjacent variances (see, e.g., Mulder et al. 2009, for testing such patterns on repeated measures means). An example of this is the hypothesis $H_2: \Sigma \in \Omega_2 = \{\Sigma : \sigma_1^2 < \sigma_2^2 < \sigma_3^2 \wedge \sigma_2^2/\sigma_1^2 < \sigma_3^2/\sigma_2^2 \wedge \Sigma > 0\}$, which states that the variance as well as the effect size $\sigma_j^2/\sigma_{j-1}^2$ increases over time. We will present an application of a test for such types of heteroscedasticity later on in the article when we apply our Bayes factor to an empirical dataset of repeated reading recognition scores in children.

Testing of variances has received much attention in the statistical literature. Numerous null hypothesis significance testing (NHST) procedures have been developed for different testing problems. Two prominent methods for testing equality of variances of independent populations are the classical likelihood ratio test based on the normal distribution and Bartlett's (1937) modification thereof. Robust alternatives to these tests were developed by Levene (1960), Brown and Forsythe (1974), and more recently by Shoemaker (2003). Noting that variances of independent populations often follow an increasing or decreasing order, Gastwirth, Gel, and Miao (2009) developed a NHST procedure for testing the null hypothesis of equal variances against an order-constrained alternative hypothesis. The problem of testing equality of variances of dependent outcomes has been considered as well. An early development is the Morgan-Pitman test (Morgan 1939; Pitman 1939) for testing equality of variances in the bivariate normal distribution. This test was modified by Wilcox (1990, 2015) to improve on the robustness of the method. Cohen (1986) and Wilcox (1989) developed extensions for testing equality of variances of more than two dependent outcomes and Mentz (2001) proposed a likelihood ratio test of equality of multiple variances in the intraclass correlation model. To our knowledge, however, no method is currently available for testing multiple hypotheses with order constraints on variances of dependent outcomes as formulated in Equation (1).

In this article, we take a Bayesian approach to the testing problem in Equation (1) using the Bayes factor (Jeffreys 1961; Kass and Raftery 1995), which is a Bayesian hypothesis testing and model selection criterion. Bayes factors have been developed for a variety of testing problems frequently encountered in practice. For example, Gönen et al. (2005) proposed a Bayesian

two-sample t -test that was later extended by Wang and Liu (2016). Klugkist, Laudy, and Hooijink (2005) developed a Bayes factor for testing order constrained hypotheses on the means in analysis of (co)variance models. Mulder (2014b) proposed Bayes factors for testing order constrained hypotheses on means and regression coefficients in the multivariate normal linear model. Gu et al. (2014) developed an approximate Bayesian procedure for testing order constrained hypotheses on regression coefficients in a structural equation modeling framework. Fosdick and Raftery (2012) used the Bayes factor to test hypotheses on the correlation in bivariate normal data. Mulder (2016) proposed Bayes factors for testing order constrained hypotheses on correlations. Recently, Böing-Messing et al. (2017) and Böing-Messing and Mulder (2018) applied the Bayes factor to the problem of testing order constrained hypotheses on the variances of independent populations. In this article, we extend this methodology by developing a Bayes factor for testing variances of dependent outcomes.

The Bayes factor has a number of advantages that are not all shared by alternative hypothesis testing approaches such as NHST by means of p -values and model selection by means of information criteria such as the AIC (Akaike 1973) and the BIC (Schwarz 1978): First, unlike Bayes factors and the AIC and BIC, p -values are not able to simultaneously test multiple hypotheses and cannot quantify evidence in favor of a hypothesis. Instead, in the NHST framework one can only test every hypothesis of interest against a common null hypothesis and the p -value is then interpreted as quantifying the evidence against this null hypothesis (although this interpretation has been contested, see, e.g. Berger and Delampady 1987). This, however, does not give us an answer as to which hypothesis receives strongest support from the data. Second, contrary to Bayes factors and the BIC, the AIC is not consistent in the sense that it is not guaranteed to select the true hypothesis as the sample size goes to infinity (e.g., O'Hagan 1995). Third, Bayes factors have an inherent Occam's razor mechanism that automatically takes the parsimony introduced by order constraints into account (e.g., Kass and Raftery 1995). While p -values do not have an inherent mechanism for measuring the complexity of a hypothesis, both the AIC and the BIC measure the complexity of a hypothesis by the number of unknown parameters. However, this is not a suitable measure when testing order constrained hypotheses. For example, under both the order constrained hypothesis $H_1: \Sigma \in \Omega_1 = \{\Sigma : \sigma_1^2 < \dots < \sigma_p^2 \wedge \Sigma > 0\}$ and the unconstrained hypothesis $H_u: \Sigma \in \Omega_u$ there are p unknown variance parameters. However, H_1 is more parsimonious than H_u because the admissible parameter space under H_1 is smaller in the sense that it is a subset of the unconstrained parameter space under H_u .

The remainder of this article is structured as follows. In the next section, we describe the multivariate normal model that is used throughout this article and we propose an unconstrained prior on the model parameters. We then develop a Bayes factor for testing order constrained hypotheses on the variances in this model using the encompassing prior approach. Following this, we examine four important aspects of the proposed Bayes factor: scale invariance, Bartlett's paradox, large sample consistency, and missing data. Subsequently, we illustrate the use of the Bayes

factor by applying it to two empirical datasets about scholastic achievements of children. We conclude the article with a discussion of our approach.

2. Model and Unconstrained Prior

In this article, we concern ourselves with testing order constrained hypotheses on the variances in the multivariate normal model

$$y_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad i = 1, \dots, n, \quad (2)$$

where n is the sample size, y_i is a p -dimensional vector of dependent variables, $\boldsymbol{\mu}_i$ denotes the mean vector for observation i , and $\boldsymbol{\Sigma}$ is an unstructured covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}. \quad (3)$$

We are interested in testing order constrained hypotheses on the variances on the main diagonal of $\boldsymbol{\Sigma}$, where the hypotheses are formulated according to Equation (1).

The mean vector, which will be treated as a nuisance parameter in this article, is typically modeled as a linear combination of a given set of covariates. For example,

$$\boldsymbol{\mu}_i = \mathbf{B}'\mathbf{x}_i \text{ under a multivariate normal linear model or} \quad (4)$$

$$\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta} \text{ under a linear mixed model,} \quad (5)$$

where \mathbf{x}_i and \mathbf{X}_i are $k \times 1$ and $p \times k$ matrices with covariates, respectively, and \mathbf{B} and $\boldsymbol{\beta}$ are $k \times p$ and $k \times 1$ matrices with regression coefficients. By writing the vectorization of the coefficients matrix in Equation (4) as $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$, we can denote the vector of nuisance parameters under both models by $\boldsymbol{\beta}$. Note that the multivariate normal linear model is a generalization of commonly used models such as (multivariate) analysis of (co)variance and multivariate regression. Furthermore, note that the standard linear mixed model, where the random effects are integrated out, has a structured covariance matrix which depends on the within-subjects variance σ^2 and the random effects covariance matrix $\boldsymbol{\Psi}$: $y_i \sim N_p(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2\mathbf{I}_p + \mathbf{X}_i\boldsymbol{\Psi}\mathbf{X}_i')$, where \mathbf{I}_p is the identity matrix of size p and the covariates for the random effects equal the covariates for the fixed effects, \mathbf{X}_i . As noted above, we use an unstructured covariance matrix instead to make sure that the model can flexibly pick up different kinds of (potentially unforeseen) covariance patterns. In Section 5.2, we will motivate this choice for the unstructured covariance matrix in a repeated measures context.

A common choice for the unconstrained prior on the covariance matrix $\boldsymbol{\Sigma}$ is a conjugate inverse Wishart prior:

$$\pi_u(\boldsymbol{\Sigma}) = W_p^{-1}(\boldsymbol{\Sigma}|\boldsymbol{\Lambda}, \nu), \quad (6)$$

where $\boldsymbol{\Lambda} > 0$ is the $p \times p$ prior scale matrix and $\nu > p - 1$ are the prior degrees of freedom. The two hyperparameters $\boldsymbol{\Lambda}$ and ν allow a flexible specification of the prior. We will come back to the choice of the hyperparameter values in Section 3.2. Since $\boldsymbol{\beta}$ contains nuisance parameters that are common under all order

constrained hypotheses, we may use the standard noninformative Jeffreys prior $\pi^N(\boldsymbol{\beta}) = C$, where C is an unspecified normalizing constant that cancels out in the computation of Bayes factors (e.g., Jeffreys 1961; O'Hagan 1995). The unconstrained joint prior on $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$ is then given by

$$\pi_u(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \pi^N(\boldsymbol{\beta}) \pi_u(\boldsymbol{\Sigma}) = C \cdot \pi_u(\boldsymbol{\Sigma}). \quad (7)$$

3. Bayes Factors for Testing Variances

3.1. The Bayes Factor

The Bayes factor for testing hypothesis H_t against a competing hypothesis $H_{t'}$ is defined as the ratio of the marginal likelihoods under the two hypotheses:

$$B_{tt'} = \frac{m_t(\mathbf{Y})}{m_{t'}(\mathbf{Y})}, \quad (8)$$

where $m_t(\mathbf{Y})$ is the marginal likelihood for observed data \mathbf{Y} under hypothesis H_t with order constraints formulated according to Equation (1). The marginal likelihood is given by

$$m_t(\mathbf{Y}) = \int_{\Omega_t} \int f_t(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \pi_t(\boldsymbol{\beta}, \boldsymbol{\Sigma}) d\boldsymbol{\beta} d\boldsymbol{\Sigma}, \quad (9)$$

where $f_t(\mathbf{Y}|\mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\Sigma})$ is the likelihood under H_t , $\pi_t(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ is the prior on $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ under H_t , and the integration region of $\boldsymbol{\Sigma}$ is restricted to the order constrained parameter space Ω_t . The marginal likelihood can be interpreted as a weighted average likelihood in the parameter space that is admissible under H_t , where the weights are given by the prior. In this way, the marginal likelihood quantifies how likely it is that the data were generated under H_t . The Bayes factor, as a ratio of marginal likelihoods, then tells us whether it is more likely that the data were generated under H_t or $H_{t'}$. If $B_{tt'} > 1$ ($B_{tt'} < 1$), then there is evidence in favor of H_t ($H_{t'}$). A Bayes factor of, say, 10 indicates that the evidence in favor of H_t is 10 times as strong as the evidence in favor of $H_{t'}$ (e.g., Jeffreys 1961).

3.2. Encompassing Prior Approach

From Equation (9), it can be seen that a prior $\pi_t(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ needs to be specified to compute the marginal likelihood under hypothesis H_t . In this article, we use the encompassing prior approach (Berger and Mortera 1999; Klugkist, Laudy, and Hooijtink 2005) to specify priors under competing order constrained hypotheses on the variances in the multivariate normal model in Equation (2). The encompassing prior approach is a popular approach to computing Bayes factors between order constrained hypotheses. The starting point is the specification of an unconstrained (or encompassing) prior under the unconstrained hypothesis H_u : $\boldsymbol{\Sigma} \in \Omega_u$. Order constrained hypotheses are nested in the unconstrained hypothesis in the sense that $\Omega_t \subset \Omega_u$. We may therefore specify priors under order constrained hypotheses as truncations of the unconstrained prior in the respective constrained subspaces. In this way, the problem of specifying a prior under each order constrained hypothesis simplifies to specifying one unconstrained prior.

The unconstrained joint prior on $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ was given in Equation (7). The prior under an order constrained hypothesis

H_t is formulated as a truncation of the unconstrained prior in the admissible parameter space Ω_t :

$$\pi_t(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \frac{1}{P(\boldsymbol{\Sigma} \in \Omega_t | H_u)} \pi_u(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma}), \quad (10)$$

where

$$P(\boldsymbol{\Sigma} \in \Omega_t | H_u) = \int_{\Omega_t} \pi_u(\boldsymbol{\Sigma}) d\boldsymbol{\Sigma} \quad (11)$$

is the prior probability that the order constraints hold under H_u , with $\pi_u(\boldsymbol{\Sigma})$ as in Equation (6), and $\mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma})$ is an indicator function that equals 1 if $\boldsymbol{\Sigma} \in \Omega_t$ and 0 otherwise. The prior probability quantifies the complexity of an order constrained hypothesis relative to the unconstrained hypothesis, where a large prior probability indicates high complexity (Mulder, Hoi-jink, and Klugkist 2010). In Equation (10), its reciprocal acts as a normalizing constant.

In this article, we consider the situation where prior information about the absolute magnitude of the (co)variances is absent, while prior information about the relative magnitude of the variances is available in the form of order constraints. Consequently, we specify a vague inverse Wishart prior under the unconstrained hypothesis. As an uninformative choice for the prior degrees of freedom we propose setting $\nu = p - 1 + 2\varepsilon$ with a small positive ε . This setting can be considered the matrix equivalent of the prior degrees of freedom in the most commonly used vague inverse gamma prior $IG(\varepsilon, \varepsilon)$ for the variance in a univariate linear regression model. To determine an uninformative choice for the prior scale matrix, first note that the unconstrained conditional posterior $\pi_u(\boldsymbol{\Sigma} | \mathbf{Y}, \boldsymbol{\beta})$ follows an inverse Wishart distribution of the form

$$W_p^{-1}(\boldsymbol{\Lambda} + (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \nu + n) \quad (12)$$

under the multivariate normal linear model, where $\mathbf{Y} = (y_1, \dots, y_n)'$ and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$, and

$$W_p^{-1}\left(\boldsymbol{\Lambda} + \sum_{i=1}^n (y_i - \mathbf{X}_i\boldsymbol{\beta})(y_i - \mathbf{X}_i\boldsymbol{\beta})', \nu + n\right) \quad (13)$$

under the linear mixed model. Consequently, we propose setting the prior scale matrix to $\boldsymbol{\Lambda} = 2\varepsilon\mathbf{I}_p$. This choice can be considered objective for two reasons: First, under this setting all possible orderings of the variances $(\sigma_1^2 < \dots < \sigma_p^2), \dots, (\sigma_p^2 < \dots < \sigma_1^2)$ are (approximately) equally likely a priori (we will come back to this in Section 4.2). As was argued by Mulder (2014a), such a specification is desirable when testing order constrained hypotheses and prior information is weak. Second, this prior scale matrix has a negligible effect on the posterior scale matrix, as can be seen from Equations (12) and (13). Thus, the posterior scale matrix is completely determined by the information in the data under both models, which implies objectivity.

Using the priors in Equations (7) and (10), it can be shown that the Bayes factor of an order constrained hypothesis H_t against the unconstrained hypothesis H_u is given by

$$B_{tu} = \frac{P(\boldsymbol{\Sigma} \in \Omega_t | \mathbf{Y}, H_u)}{P(\boldsymbol{\Sigma} \in \Omega_t | H_u)}, \quad (14)$$

where

$$P(\boldsymbol{\Sigma} \in \Omega_t | \mathbf{Y}, H_u) = \int_{\Omega_t} \pi_u(\boldsymbol{\Sigma} | \mathbf{Y}) d\boldsymbol{\Sigma} \quad (15)$$

is the posterior probability that the order constraints hold under H_u . A proof is given in Appendix A. The posterior probability quantifies the fit of an order constrained hypothesis relative to the unconstrained hypothesis, where a large posterior probability indicates a good fit. From Equation (14), it can be seen that the proposed Bayes factor automatically functions as an Occam's razor by taking the fit and the complexity of order constrained hypotheses into account. The Bayes factor between two order constrained hypotheses H_t and $H_{t'}$ is given by

$$B_{tt'} = \frac{B_{tu}}{B_{t'u}}. \quad (16)$$

Note that the posterior probability of an order constrained hypothesis H_t can be computed using the Bayes factors against the unconstrained hypothesis according to

$$P(H_t | \mathbf{Y}) = \frac{B_{tu}P(H_t)}{\sum_{t'=1}^T B_{t'u}P(H_{t'})}, \quad (17)$$

where $P(H_t)$ is the prior probability of H_t . If prior information about the hypotheses is weak, then it is customary to specify equal prior probabilities $P(H_1) = \dots = P(H_T) = 1/T$. From Equations (11) and (14)–(17), it can be seen that the Bayes factors and posterior probabilities of the hypotheses only depend on the unconstrained prior and posterior of $\boldsymbol{\Sigma}$. In Appendix B, we describe a method for computing the Bayes factors.

4. Properties of the Bayes Factor for Testing Variances

4.1. Scale Invariance

Scale invariance implies that rescaling the data \mathbf{Y} by multiplying all elements by a constant w should not change the evidence in favor of competing hypotheses as indicated by the Bayes factor because the relative variability remains unchanged. Our Bayes factor for testing variances is in fact scale invariant. For the multivariate normal linear model, this can be shown as follows: First, note that under this model the unconstrained marginal posterior of $\boldsymbol{\Sigma}$ given the original data \mathbf{Y} follows an $W_p^{-1}(2\varepsilon\mathbf{I}_p + \mathbf{S}, \nu + n - k)$ distribution, where $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ is the residual sums of squares matrix with $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. For the rescaled data $w\mathbf{Y}$, the residual sums of squares matrix is $w^2\mathbf{S}$. The posterior scale matrix is then equal to $2\varepsilon\mathbf{I}_p + w^2\mathbf{S} \approx w^2\mathbf{S}$. Since the prior scale matrix can be ignored, the posterior probability that the order constraints on the variances hold can be written as

$$\begin{aligned} P(\boldsymbol{\Sigma} \in \Omega_t | w\mathbf{Y}, H_u) &= \int_{\boldsymbol{\Sigma} \in \Omega_t} W_p^{-1}(\boldsymbol{\Sigma} | w^2\mathbf{S}, \nu + n - k) d\boldsymbol{\Sigma} \\ &= \int_{w^2\boldsymbol{\Phi} \in \Omega_t} W_p^{-1}(\boldsymbol{\Phi} | \mathbf{S}, \nu + n - k) d\boldsymbol{\Phi} \\ &= \int_{\boldsymbol{\Phi} \in \Omega_t} W_p^{-1}(\boldsymbol{\Phi} | \mathbf{S}, \nu + n - k) d\boldsymbol{\Phi} \\ &= P(\boldsymbol{\Sigma} \in \Omega_t | \mathbf{Y}, H_u), \end{aligned} \quad (18)$$

where in the first step we applied the transformation $\boldsymbol{\Phi} = w^{-2}\boldsymbol{\Sigma}$. Furthermore, order constrained subspaces are considered such that $\boldsymbol{\Phi} \in \Omega_t$ if $w^2\boldsymbol{\Phi} \in \Omega_t$, which explains the second step. As a result, our Bayes factor is invariant to the scale of the data. The same argument can be used for the linear mixed model.

4.2. Bartlett's Paradox

Bartlett's paradox (Bartlett 1957; Lindley 1957; Jeffreys 1961) is the situation where a Bayes factor for testing a precise null hypothesis against an unconstrained alternative hypothesis favors the null regardless of the information in the data if the variance of the prior under the alternative hypothesis is specified sufficiently large. Typically, Bartlett's paradox is not an issue in the case of testing order constrained hypotheses (Klugkist and Hoijtink 2007; Mulder 2014a). To see this in our setup, we need to investigate whether the prior probability that the order constraints hold under H_u , that is, the denominator of the Bayes factor expression in Equation (14), can be made arbitrarily small by specifying ε in the unconstrained prior $W_p^{-1}(2\varepsilon\mathbf{I}_p, p - 1 + 2\varepsilon)$ small enough. When considering a specific ordering of the variances, say, $\sigma_1^2 < \dots < \sigma_p^2$, thorough numerical checks (setting $\varepsilon = 0.1, 0.01, 0.001$, etc.) revealed that the prior probability is virtually independent of the exact choice of ε and is extremely close to $1/p!$. Although this finding satisfies our intuition (as there are $p!$ possible orderings of p variances), it cannot be theoretically proven because the variances are not independent and the marginal prior for the variances does not have a known distribution. Also for other order constrained variance patterns the prior probability is virtually independent of the exact choice of a small ε . Furthermore, since the posterior scale matrix for Σ depends on the sum of the prior scale $2\varepsilon\mathbf{I}_p$ and the sums of squares in the data, the posterior probability that the order constraints hold is virtually independent of ε as well. Thus, we can conclude that the Bayes factor is practically independent of the exact choice of $\varepsilon > 0$ and Bartlett's paradox is not an issue for our order constrained tests on variances.

4.3. Large Sample Consistency

Bayes factors are large sample consistent under very general conditions, which means that the posterior probability of the true hypothesis tends to 1 as the sample size tends to infinity. For our Bayes factor, this important property can be explained as follows: Suppose that the true values of the variances lie in the admissible parameter space Ω_t of hypothesis H_t . Then, as the sample size increases, the unconstrained posterior of Σ concentrates in Ω_t , so that the probability $P(\Sigma \in \Omega_t | Y, H_u)$ tends to 1. This implies that if $H_{t'}$ with admissible parameter space $\Omega_{t'}$ is a competing hypothesis that does not overlap with H_t (in the sense that $\Omega_t \cap \Omega_{t'} = \emptyset$), then $P(\Sigma \in \Omega_{t'} | Y, H_u)$ tends to 0. From Equation (16), it can be seen that the Bayes factor of H_t against $H_{t'}$ tends to infinity in this case (note that the prior probabilities $P(\Sigma \in \Omega_t | H_u)$ and $P(\Sigma \in \Omega_{t'} | H_u)$ are constants that do not depend on the data). As a result, the posterior probability of H_t tends to 1, which implies that the Bayes factor in Equation (16) is consistent. The same argument applies in the case where there are multiple competing hypotheses that do not overlap with H_t . To ensure that the Bayes factor behaves consistently it is advisable to specify the hypotheses such that they do not overlap and cover the entire unconstrained parameter space (in the sense that $\Omega_1 \cup \dots \cup \Omega_T = \Omega_u$).

4.4. Hypothesis Testing in the Case of Missing Data

As shown by Hoijtink et al. (2019), missing observations that are missing at random can be handled via multiple imputation when computing Bayes factors. This imputation step can be incorporated in the posterior sampler in a natural manner in the Bayesian framework. Note that multiple imputation is the preferred method over listwise deletion, which results in a loss of information and possibly bias (Rubin 1987, 1996). Below, we show how the Bayes factors for testing order constraints on variances of dependent outcomes are obtained in the case of missing observations.

As was shown in Section 3.2, only the posterior probability that the order constraints hold under the unconstrained hypothesis needs to be computed to obtain the Bayes factors for the order constrained test. Therefore, we do not need separate missing data models under the different order constrained hypotheses but only one missing data model specified under the unconstrained hypothesis. Let us denote the missing data by \mathbf{Z} and the observed data by \mathbf{Y} . In the Bayesian framework, the missing data \mathbf{Z} are naturally treated as unknown parameters just like the model parameters. Inferences are then based on the joint posterior distribution $\pi_u(\boldsymbol{\beta}, \Sigma, \boldsymbol{\zeta}, \mathbf{Z} | \mathbf{Y})$ of the model parameters and the missing data, where $\boldsymbol{\zeta}$ are additional parameters used for the missing data model. The marginal posterior distribution of Σ can be obtained by integrating the joint posterior distribution over the model parameters $\boldsymbol{\beta}, \boldsymbol{\zeta}$ and the missing data \mathbf{Z} . The posterior probability that the order constraints hold can then be computed as

$$\begin{aligned} P(\Sigma \in \Omega_t | \mathbf{Y}, H_u) &= \int_{\Omega_t} \pi_u(\Sigma | \mathbf{Y}) d\Sigma \\ &= \int_{\Omega_t} \iiint \pi_u(\boldsymbol{\beta}, \Sigma, \boldsymbol{\zeta}, \mathbf{Z} | \mathbf{Y}) d\boldsymbol{\beta} d\boldsymbol{\zeta} d\mathbf{Z} d\Sigma \\ &\approx \frac{1}{M} \sum_{m=1}^M \mathbf{1}_{\Omega_t}(\boldsymbol{\Sigma}^{(m)}), \end{aligned} \tag{19}$$

where $\boldsymbol{\Sigma}^{(1)}, \dots, \boldsymbol{\Sigma}^{(M)}$ are posterior draws for the covariance matrix. Thus, we can approximate the posterior probability using a Gibbs sampling scheme in which we iteratively draw samples from the full conditional distributions of $\boldsymbol{\beta}, \Sigma, \boldsymbol{\zeta}$, and \mathbf{Z} . This approach is also referred to as data augmentation because the observed data are augmented with simulated values from the posterior predictive distribution of the missing data \mathbf{Z} to obtain a complete dataset (e.g., Tanner and Wong 1987; Gelman et al. 2013, chap. 18). When computing the posterior probability in Equation (19) this way, the uncertainty caused by the missing data is integrated out as is usual in a Gibbs sampler. Note that this is only allowed because the missing data are missing at random, which assumes that missingness is conditionally independent of the missing data given the observed data. In Section 5.2, we will present an analysis of an empirical dataset with missing values where we make use of the missing data approach described here.

5. Empirical Applications

5.1. Heteroscedasticity in Cross-Sectional Data: Peabody Individual Achievement Test

The Peabody Individual Achievement Test (PIAT; Dunn and Markwardt 1970) is a survey for measuring an individual's scholastic attainment in the areas of general information, reading, writing, and mathematics. As part of the 1979 National Longitudinal Survey of Youth (Baker et al. 1993), a total of 997 children completed three PIAT subtests at age 10: Mathematics, Reading Recognition, and Reading Comprehension. All subtests consisted of 84 multiple choice questions and the final score on each subtest was the number of correct answers. Our interest was in whether mathematics scores were less or more variable than reading scores. Consequently, we tested the following competing hypotheses on the variances of the subtests:

$$\begin{aligned} H_1: \Sigma \in \Omega_1 &= \{ \Sigma : \sigma_M^2 < (\sigma_R^2, \sigma_C^2) \wedge \Sigma > 0 \}, \\ H_2: \Sigma \in \Omega_2 &= \{ \Sigma : (\sigma_R^2, \sigma_C^2) < \sigma_M^2 \wedge \Sigma > 0 \}, \\ H_3: \neg(H_1 \vee H_2), \end{aligned} \quad (20)$$

where σ_M^2 , σ_R^2 , and σ_C^2 denote the variance of the scores on the Mathematics, Reading Recognition, and Reading Comprehension subtests, respectively. Note that the comma between σ_R^2 and σ_C^2 means that there was no order constraint on the variances of the two reading subtests. We tested H_1 and H_2 against their complement H_3 to take into account that the data might support neither H_1 nor H_2 .

We fitted a multivariate normal model without covariates to the data from the three PIAT subtests. There were no missing values in the dataset. We specified the hyperparameters of the unconstrained prior on the covariance matrix as $\Lambda = 0.002 \cdot I_3$ and $\nu = 2.002$. Subsequently, we performed a sequential analysis of the data, where we recomputed the Bayes factors and posterior probabilities of the hypotheses for every additional 10 observations (i.e., we conducted the hypothesis test on the first 10, 20, 30, ... observations in the dataset). We did so to illustrate how the evidence in favor of the hypotheses changes as more data become available. In practice, such a sequential approach allows researchers to update the evidence in favor of the hypotheses during data collection. Depending on the amount of evidence at a given point in time, it can be decided whether more data are needed to draw a more decisive conclusion about which hypothesis is most likely to be true. Note that Bayes factors are well-suited for such a sequential analysis because they are not affected by Type I error inflation and hence multiple testing is not a problem (e.g., Berger and Mortera 1999).

The results of our sequential analysis are shown in Figures 1(a) and (b). To enhance readability, Figure 1(a) shows the logarithm of the Bayes factor, which is also referred to as the weight of evidence (WOE): $\text{WOE}_{it'} = \log(B_{it'})$. The line for WOE_{12} is discontinued because for sample sizes greater than 600 the posterior probability that the order constraints of H_2 hold was numerically approximated as 0 (see Appendix B), which resulted in an infinite weight of evidence. Figure 1(b) shows the posterior probabilities of the hypotheses, which were computed assuming equal prior probabilities. From the weights of evidence and the posterior probabilities, it can be seen that H_2

received strongest support from the data for sample sizes up to and including 70. After that, H_1 received strongest support, and there was a trend for the evidence in favor of H_1 to increase as more data arrived. In Figure 1(b), the lines for H_1 and H_3 behave inversely for sample sizes greater than 200 because H_2 receives no support here so that an increase for one of H_1 and H_3 is necessarily accompanied by a decrease for the other. The results show that for the final sample size of 997, we can practically rule out H_2 and H_3 and conclude that mathematics scores were less variable than reading scores, as is stated by H_1 .

5.2. Heteroscedasticity in Repeated Measures Data: Reading Recognition

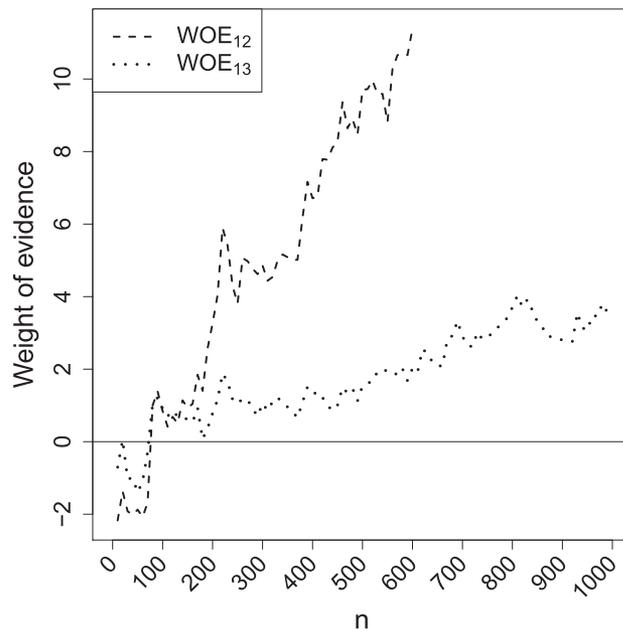
An important topic in repeated measures analysis is heteroscedasticity, which focuses on the variability of observations over time. It is often of interest whether a destabilizing effect is present, that is, whether the variability increases over time. This will also be the focus here.

A popular model for analyzing repeated measures data that allows for heteroscedastic variances is the mixed effects model with a random intercept and a random slope for time (e.g., Hedeker and Gibbons 2006, chap. 4). The advantage of such a mixed model is that it is capable of separately modeling the variability at the within-subjects level and the between-subjects level. On the other hand, the random slope model has some drawbacks for our purpose of testing order constrained hypotheses on the variances of dependent outcomes: The model implies a specific pattern of the variances and covariances in the marginal model, namely, $\Sigma = \sigma^2 I_p + X_i \Psi X_i'$. In the case of the popular random intercept and random slope model for balanced data with four repeated measurements at time $t = (t_1, \dots, t_4)'$, the covariance matrix equals

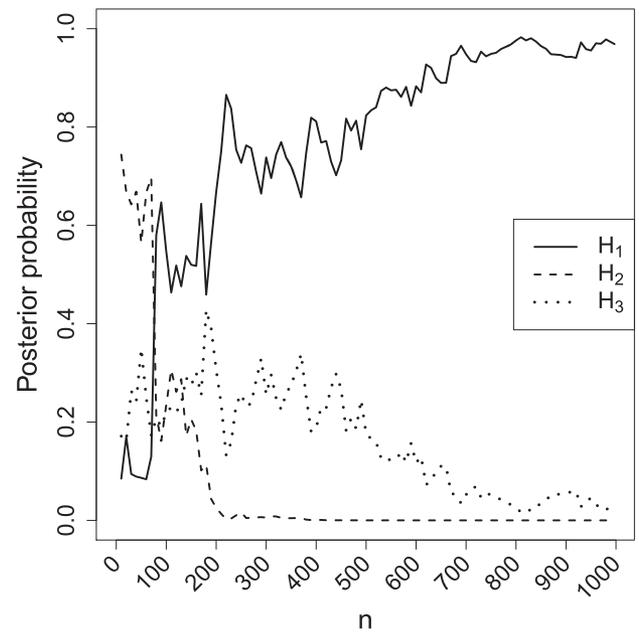
$$\begin{aligned} \Sigma &= \sigma^2 I_4 + [\mathbf{1}_4 \quad \mathbf{t}] \Psi [\mathbf{1}_4 \quad \mathbf{t}]' \\ &= \sigma^2 I_4 + \mathbf{1}_4 \mathbf{1}_4' \psi_1^2 + (\mathbf{1}_4 \mathbf{t}' + \mathbf{t} \mathbf{1}_4') \psi_{12} + \mathbf{t} \mathbf{t}' \psi_2^2, \end{aligned} \quad (21)$$

where $\mathbf{1}_4 = (1, 1, 1, 1)'$. Since $t_1 < \dots < t_4$, the (co)variances follow a very specific pattern which highly depends on whether the time points are centered around 0 (as is typically the case). Also observe that the variability of the observations over time highly depends on whether ψ_2^2 is (approximately) zero or not. Testing such boundary values can be quite complex however (Mulder and Fox 2019). For this reason, we consider the unstructured covariance matrix as a flexible tool for testing competing patterns of heteroscedasticity in repeated measures data, as we show below. Note that this model assumes balanced data with equal measurement occasions across subjects. Further note that by testing variance patterns under an unstructured covariance matrix, the proposed method can also be used when building linear mixed models to check if specific variance patterns implied by a linear mixed model are actually supported by the data. We will address another interesting choice for the covariance structure in the conclusion of this article in Section 6.

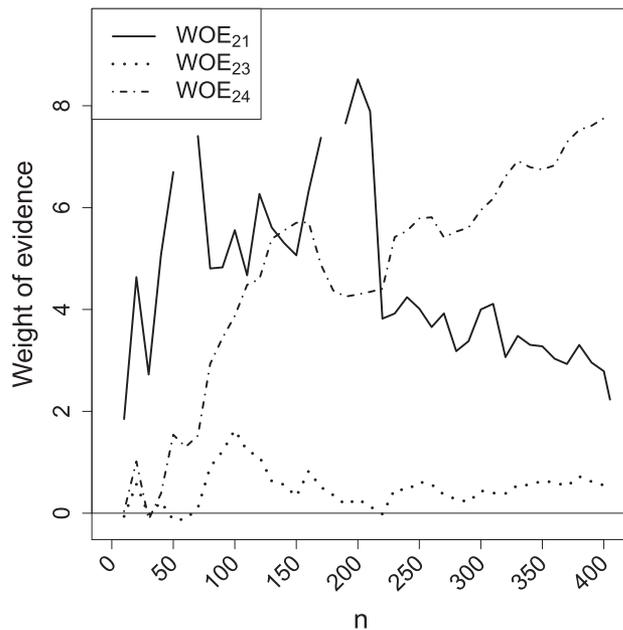
The repeated measures application we present in this section is based on an empirical dataset discussed in Vermunt and Magidson (2005). The outcome of interest in this dataset was reading recognition in children as measured by the Reading Recognition subtest of the PIAT. The Reading Recognition



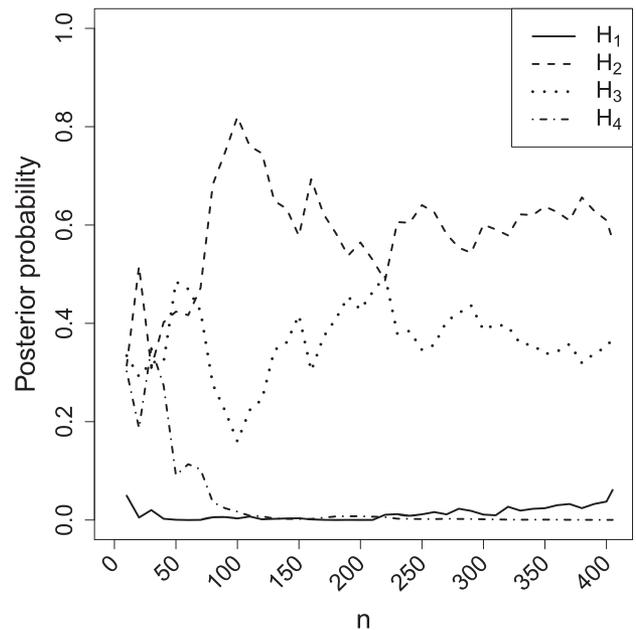
(a) Peabody Individual Achievement Test, weight of evidence (WOE)



(b) Peabody Individual Achievement Test, posterior probabilities



(c) Reading recognition, weight of evidence (WOE)



(d) Reading recognition, posterior probabilities

Figure 1. Results of sequential analyses of two empirical datasets. (a, b) The weight of evidence (where $WOE_{tr} = \log(B_{tr})$) and the posterior probabilities of the hypotheses (assuming equal prior probabilities), respectively, for the Peabody Individual Achievement Test data discussed in Section 5.1. The weights of evidence and posterior probabilities were computed on the first 10, 20, 30, ... observations in the dataset. (c, d) The corresponding results for the reading recognition data discussed in Section 5.2. See the text for a description of the hypotheses that were tested in each of the two sequential analyses. In (a) and (c), the lines for WOE_{12} and WOE_{21} are interrupted due to numerical reasons (see text for details).

subtest measures a child's word recognition and pronunciation ability. The subtest was administered four times at two-year intervals to a sample of 405 children. We tested the following hypotheses on the variances of the repeated reading recognition measures and the ratios of adjacent variances:

H_1 : increasing variances and increasing ratios of adjacent variances over time,

H_2 : increasing variances and decreasing ratios of adjacent variances over time,

$$\begin{aligned}
H_3: & \text{ increasing variances and neither increasing nor} \\
& \text{ decreasing ratios over time,} \\
H_4: & \text{ not } H_1, H_2, \text{ or } H_3.
\end{aligned} \tag{22}$$

These hypotheses can be expressed formally using order constraints on the variances and ratios of adjacent variances as follows:

$$H_1: \Sigma \in \Omega_1 = \{ \Sigma : \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2 \wedge \sigma_2^2/\sigma_1^2 < \sigma_3^2/\sigma_2^2 < \sigma_4^2/\sigma_3^2 \wedge \Sigma > 0 \},$$

$$H_2: \Sigma \in \Omega_2 = \{ \Sigma : \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2 \wedge \sigma_4^2/\sigma_3^2 < \sigma_3^2/\sigma_2^2 < \sigma_2^2/\sigma_1^2 \wedge \Sigma > 0 \},$$

$$\begin{aligned}
H_3: \Sigma \in \Omega_3 = & \{ \Sigma : \sigma_1^2 < \sigma_2^2 < \sigma_3^2 < \sigma_4^2 \wedge \\
& (\neg(\sigma_2^2/\sigma_1^2 < \sigma_3^2/\sigma_2^2 < \sigma_4^2/\sigma_3^2 \vee \sigma_4^2/\sigma_3^2 \\
& < \sigma_3^2/\sigma_2^2 < \sigma_2^2/\sigma_1^2)) \wedge \Sigma > 0 \},
\end{aligned}$$

$$H_4: \neg(H_1 \vee H_2 \vee H_3), \tag{23}$$

where σ_j^2 is the variance of the j th measurement. Thus, H_1 , H_2 , and H_3 all state that the variability of reading recognition performances increased over time. Such an increase in variability is frequently observed in practice (e.g., Aunola et al. 2004; Böing-Messing et al. 2017; Hedeker and Gibbons 2006). Moreover, H_1 (H_2) states that the effect size $\sigma_j^2/\sigma_{j-1}^2$ increased (decreased) over time, whereas according to H_3 the effect size neither steadily increased nor decreased over time. Similar as in the previous section, we tested H_1 , H_2 , and H_3 against their complement H_4 to safeguard against the data supporting neither of the three hypotheses.

We fitted a multivariate normal model with a quadratic effect of (centered) time and an unstructured covariance matrix to the repeated reading recognition data: $y_i \sim N_4([\mathbf{1}_4 \quad \mathbf{t} \quad \mathbf{t}^2]\boldsymbol{\beta}, \Sigma)$, where $\mathbf{t} = (-1.5, -0.5, 0.5, 1.5)'$ and $\mathbf{t}^2 = (2.25, 0.25, 0.25, 2.25)'$. The dataset contained missing values on the outcome: In total, 18.21% of the repeated reading recognition values were missing. Hence, we used the imputation approach described in Section 4.4 to compute the Bayes factors. Since the missingness only occurred on the outcome, there was no need to specify an extended missing data model. Instead, we used the observation model above to impute the missing values in the Gibbs sampler. Again, we specified the unconstrained prior on the covariance matrix to be uninformative by setting the hyperparameters equal to $\Lambda = 0.002 \cdot \mathbf{I}_4$ and $\nu = 3.002$. As in the previous section, we performed a sequential analysis of the data by testing the hypotheses on the first 10, 20, 30, ... observations in the dataset.

Figures 1(c) and (d) present the weights of evidence and the posterior probabilities of the hypotheses resulting from the sequential analysis, respectively. As before, the posterior probabilities were computed assuming equal prior probabilities of the hypotheses. In Figure 1(c), the line for WOE_{21} is interrupted because for sample sizes of 60 and 180 the posterior probability that the order constraints of H_1 hold was approximated as 0. The results show that the evidence fluctuated for sample sizes up to 60, with H_2 , H_3 , and H_4 receiving strongest support from the data at some point. For sample sizes greater than 60, H_2 received strongest support except when the sample size was

220, for which H_3 was most supported. In fact, it can be seen that H_3 remained a viable competitor throughout the sequential analysis. The results indicate that the complement H_4 can be practically ruled out, whereas H_1 received weak support at the end of the sequential analysis. Based on these results, we may conclude that there was strong evidence that the variance increased over time. Moreover, while there was considerable evidence that the effect size (as measured by the ratio of adjacent variances) decreased over time (as stated by H_2), it cannot be ruled out that the effect size developed in a different way over time (as stated by H_3 and H_1).

6. Conclusion

In this article, we developed a Bayes factor for testing order constrained hypotheses on the variances of dependent outcomes. We applied the encompassing prior approach, in which priors under order constrained hypotheses are formulated as truncations of the prior under the unconstrained hypothesis. This approach to prior specification has two main advantages: First, the problem of specifying a prior under every order constrained hypothesis to be tested simplifies to specifying one unconstrained prior. This is a useful property in practice, since in our experience researchers find it difficult to formulate subjective priors on (co)variances, in particular under order constrained hypotheses. We used a conjugate inverse Wishart distribution with noninformative hyperparameters as the unconstrained prior on the covariance matrix. With this specification, the Bayes factor is fully automatic in the sense that there is no need to specify a subjective prior under any hypothesis. The second advantage of the encompassing prior approach is that Bayes factors between order constrained hypotheses can be computed straightforwardly by sampling from the prior and the posterior distribution of the covariance matrix and computing the proportion of draws that satisfy the order constraints. The Bayes factor is consistent in that it is guaranteed to select the true hypothesis as the sample size increases.

A natural extension of the testing problem we considered in this article is to enable equality constraints between the variances (e.g., $H_1: \Sigma \in \Omega_1 = \{ \Sigma : \sigma_1^2 = \sigma_2^2 < \sigma_3^2 \wedge \Sigma > 0 \}$). This testing problem entails two main challenges: First, one needs to ensure positive definiteness of the covariance matrix under equality and inequality constrained hypotheses. A useful method to achieve this could be the separation strategy of Barnard, McCulloch, and Meng (2000). In this approach, the covariance matrix is decomposed as $\Sigma = \mathbf{D}\mathbf{C}\mathbf{D}$, where \mathbf{D} is a diagonal matrix of standard deviations and \mathbf{C} is a positive definite correlation matrix. Equality and inequality constrained hypotheses could then be formulated on the standard deviations in \mathbf{D} . Second, Bartlett's paradox will become a problem if priors under composite hypotheses are specified too vaguely. A possible solution to this problem is to let priors under composite hypotheses contain the information of a minimal experiment (e.g., Berger and Pericchi 1996; Mulder, Hoijtink, and de Leeuw 2012; Böing-Messing and Mulder 2018). This can be achieved by updating an uninformative improper prior with the information contained in the smallest possible subset of the sample data that results in a proper posterior distribution. The remaining part

of the sample data is then used to compute the Bayes factors between the (in)equality constrained hypotheses under investigation. Based on our findings in this article, we expect that the Bayes factor is well-suited for the intricate problem of testing multiple hypotheses with equality and inequality constraints on the variances.

Another interesting direction for future research would be to test constraints on variances under different covariance structures. To get a better understanding of where the variability in the data comes from, the implied covariance matrix in the mixed effects model, $\Sigma = \sigma^2 I_p + X_i \Psi X_i'$, can be made less restrictive by allowing the within-subjects variances to vary over time, that is, $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_p^2) + X_i \Psi X_i'$. This will provide insight into whether the variability changes as a result of within-subjects variability, as a result of between-subjects variability, or as a result of both (see also Mulder and Fox 2019 for a Bayes factor test on between-subjects variances). It is then important that the structured part of the covariance matrix, $X_i \Psi X_i'$, fits the data well because a misfit would distort the estimates of the diagonal elements $\sigma_1^2, \dots, \sigma_p^2$. The proposed Bayes factors can also be used when building other types of mixed effects models, structural equation models, or time series models, by testing whether specific order constrained patterns of the variance components are supported by the data.

Appendix A: Bayes Factor of H_t Against H_u

First, note that the likelihood under an order constrained hypothesis H_t is a truncation of the unconstrained likelihood, that is, $f_t(Y|X, \beta, \Sigma) = f_u(Y|X, \beta, \Sigma) \mathbf{1}_{\Omega_t}(\Sigma)$, where $\mathbf{1}_{\Omega_t}(\Sigma)$ is an indicator function that equals 1 if $\Sigma \in \Omega_t$ and 0 otherwise. The Bayes factor of H_t against the unconstrained hypothesis H_u is given by

$$\begin{aligned} B_{tu} &= \frac{m_t(Y)}{m_u(Y)} = \frac{\int_{\Omega_t} \int f_t(Y|X, \beta, \Sigma) \pi_t(\beta, \Sigma) d\beta d\Sigma}{\int_{\Omega_u} \int f_u(Y|X, \beta, \Sigma) \pi_u(\beta, \Sigma) d\beta d\Sigma} \\ &= \frac{\int_{\Omega_t} \int f_u(Y|X, \beta, \Sigma) P(\Sigma \in \Omega_t | H_u)^{-1} \pi_u(\beta, \Sigma) \mathbf{1}_{\Omega_t}(\Sigma) d\beta d\Sigma}{\int_{\Omega_u} \int f_u(Y|X, \beta, \Sigma) \pi_u(\beta, \Sigma) d\beta d\Sigma} \\ &= \frac{1}{P(\Sigma \in \Omega_t | H_u)} \int_{\Omega_t} \int \frac{f_u(Y|X, \beta, \Sigma) \pi_u(\beta, \Sigma)}{\int_{\Omega_u} \int f_u(Y|X, \beta, \Sigma) \pi_u(\beta, \Sigma) d\beta d\Sigma} \\ &\quad \times d\beta d\Sigma \\ &= \frac{1}{P(\Sigma \in \Omega_t | H_u)} \int_{\Omega_t} \int \pi_u(\beta, \Sigma | Y) d\beta d\Sigma \\ &= \frac{1}{P(\Sigma \in \Omega_t | H_u)} \int_{\Omega_t} \pi_u(\Sigma | Y) d\Sigma = \frac{P(\Sigma \in \Omega_t | Y, H_u)}{P(\Sigma \in \Omega_t | H_u)}, \end{aligned} \tag{A.1}$$

where in the third line we may omit the indicator function because the integration region is already restricted to the constrained parameter space Ω_t .

Appendix B: Computation of the Bayes Factor

To compute the Bayes factors and posterior probabilities of the hypotheses, we need to evaluate the integrals in Equations (11) and (15). This cannot be done analytically, but the integrals can be approximated numerically by drawing a large sample from the respective unconstrained prior and posterior distribution of Σ and then computing the proportion of draws that fall in the admissible parameter space Ω_t : To compute the prior probability in Equation (11) under

both models, we first sample $\Sigma^{(m)} \sim W_p^{-1}(\Lambda, \nu)$, for $m = 1, \dots, M$, where M is the number of draws and $\Sigma^{(m)}$ is the m th draw. We can then approximate the prior probability that the order constraints hold by computing the proportion of draws from the prior that satisfy the constraints, that is, $P(\Sigma \in \Omega_t | Y, H_u) \approx \frac{1}{M} \sum_{m=1}^M \mathbf{1}_{\Omega_t}(\Sigma^{(m)})$. To compute the posterior probability in Equation (15) under the multivariate normal linear model, first note that under this model the unconstrained marginal posterior of Σ follows an $W_p^{-1}(\Lambda + S, \nu + n - k)$ distribution, where $S = (Y - X\hat{B})'(Y - X\hat{B})$ is the residual sums of squares matrix with $Y = (y_1, \dots, y_n)'$, $X = (x_1, \dots, x_n)'$, and $\hat{B} = (X'X)^{-1}X'Y$. Hence, we can use the above sampling approach to approximate the posterior probability that the order constraints hold. To obtain draws from the unconstrained posterior of Σ under the linear mixed model, a Gibbs sampler can be used. Here, one iteratively samples from the conditional posterior of Σ given β in Equation (13) and the conditional posterior of β given Σ , which has a multivariate normal distribution $N_k((X'\Sigma_n^{-1}X)^{-1}X'\Sigma_n^{-1}y, (X'\Sigma_n^{-1}X)^{-1})$, where $y = (y'_1, \dots, y'_n)'$, $X = (X'_1, \dots, X'_n)'$, $\Sigma_n = I_n \otimes \Sigma$, and \otimes denotes the Kronecker product.

A Monte Carlo error of the Bayes factor estimates can be quantified using the method of Mulder, Hoijtink, and de Leeuw (2012). Let us refer to the posterior draws that satisfy the order constraints as “successes” and the draws that do not satisfy the order constraints as “failures” in a binomial experiment where the unknown posterior probability is the success probability and M is the number of trials. When setting a uniform prior on the unknown posterior probability, its posterior follows a Beta($\nu + 1, M - \nu + 1$) distribution, where ν is the number of draws satisfying the constraints. Subsequently, posterior draws can be generated from this distribution to compute realizations of the Bayes factor. The standard deviation of these realizations is then the Monte Carlo error of the estimated Bayes factor.

Acknowledgments

The authors wish to thank the editor, an anonymous associate editor, three anonymous reviewers, and Marcel van Assen for valuable comments on earlier versions of this article. The authors thank Jeroen Vermunt for providing the data for the second empirical application.

Funding

This research was partly supported by a Rubicon grant which was awarded to the second author by the Netherlands Organization for Scientific Research (NWO).

References

- Akaike, H. (1973), “Information Theory and an Extension of the Maximum Likelihood Principle,” in *2nd International Symposium on Information Theory*, eds. B. N. Petrov and F. Csáki, Budapest, Hungary: Akadémiai Kiadó, pp. 267–281. [2]
- Aunola, K., Leskinen, E., Lerkkanen, M.-K., and Nurmi, J.-E. (2004), “Developmental Dynamics of Math Performance From Preschool to Grade 2,” *Journal of Educational Psychology*, 96, 699–713. [1,8]
- Baker, P. C., Keck, C. K., Mott, F. L., and Quinlan, S. V. (1993), *NLSY Child Handbook: A Guide to the 1986–1990 National Longitudinal Survey of Youth Child Data* (Revised ed.), Columbus, OH: The Ohio State University. [6]
- Barnard, J., McCulloch, R., and Meng, X.-L. (2000), “Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, With Application to Shrinkage,” *Statistica Sinica*, 10, 1281–1311. [8]
- Bartlett, M. S. (1937), “Properties of Sufficiency and Statistical Tests,” *Proceedings of the Royal Society of London, Series A*, 160, 268–282. [2]

- (1957), “A Comment on D. V. Lindley’s Statistical Paradox,” *Biometrika*, 44, 533–534. [5]
- Berger, J. O., and Delampady, M. (1987), “Testing Precise Hypotheses,” *Statistical Science*, 2, 317–335. [2]
- Berger, J. O., and Mortera, J. (1999), “Default Bayes Factors for Nonnested Hypothesis Testing,” *Journal of the American Statistical Association*, 94, 542–554. [3,6]
- Berger, J. O., and Pericchi, L. R. (1996), “The Intrinsic Bayes Factor for Model Selection and Prediction,” *Journal of the American Statistical Association*, 91, 109–122. [8]
- Böing-Messing, F., and Mulder, J. (2016), “Automatic Bayes Factors for Testing Variances of Two Independent Normal Distributions,” *Journal of Mathematical Psychology*, 72, 158–170. [1]
- (2018), “Automatic Bayes Factors for Testing Equality- and Inequality-Constrained Hypotheses on Variances,” *Psychometrika*, 83, 586–617. [1,2,8]
- Böing-Messing, F., van Assen, M. A. L. M., Hofman, A. D., Hoijsink, H., and Mulder, J. (2017), “Bayesian Evaluation of Constrained Hypotheses on Variances of Multiple Independent Groups,” *Psychological Methods*, 22, 262–287. [1,2,8]
- Brown, M. B., and Forsythe, A. B. (1974), “Robust Tests for the Equality of Variances,” *Journal of the American Statistical Association*, 69, 364–367. [2]
- Carroll, R. J. (2003), “Variances Are Not Always Nuisance Parameters,” *Biometrics*, 59, 211–220. [1]
- Cohen, A. (1986), “Comparing Variances of Correlated Variables,” *Psychometrika*, 51, 379–391. [2]
- Dunn, L. M., and Markwardt, F. C. (1970), *Peabody Individual Achievement Test*, Circle Pines, MN: American Guidance Service. [6]
- Fosdick, B. K., and Raftery, A. E. (2012), “Estimating the Correlation in Bivariate Normal Data With Known Variances and Small Sample Sizes,” *The American Statistician*, 66, 34–41. [2]
- Gastwirth, J. L., Gel, Y. R., and Miao, W. (2009), “The Impact of Levene’s Test of Equality of Variances on Statistical Theory and Practice,” *Statistical Science*, 24, 343–360. [2]
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian Data Analysis* (3rd ed.), Boca Raton, FL: Chapman and Hall/CRC. [5]
- Gönen, M., Johnson, W. O., Lu, Y., and Westfall, P. H. (2005), “The Bayesian Two-Sample t Test,” *The American Statistician*, 59, 252–257. [2]
- Gu, X., Mulder, J., Deković, M., and Hoijsink, H. (2014), “Bayesian Evaluation of Inequality Constrained Hypotheses,” *Psychological Methods*, 19, 511–527. [2]
- Hedeker, D., and Gibbons, R. D. (2006), *Longitudinal Data Analysis*, Hoboken, NJ: Wiley. [1,6,8]
- Hoijsink, H., Gu, X., Mulder, J., and Rosseel, Y. (2019), “Computing Bayes Factors From Data With Missing Values,” *Psychological Methods*, 24, 253–268. [5]
- Hultsch, D. F., MacDonald, S. W. S., and Dixon, R. A. (2002), “Variability in Reaction Time Performance of Younger and Older Adults,” *Journal of Gerontology: Psychological Sciences*, 57B, 101–115. [1]
- Jeffreys, H. (1961), *Theory of Probability* (3rd ed.), Oxford: Oxford University Press. [2,3,5]
- Kaptein, M., and Eckles, D. (2012), “Heterogeneity in the Effects of Online Persuasion,” *Journal of Interactive Marketing*, 26, 176–188. [1]
- Kass, R. E., and Raftery, A. E. (1995), “Bayes Factors,” *Journal of the American Statistical Association*, 90, 773–795. [2]
- Klugkist, I., and Hoijsink, H. (2007), “The Bayes Factor for Inequality and About Equality Constrained Models,” *Computational Statistics and Data Analysis*, 51, 6367–6379. [5]
- Klugkist, I., Laudy, O., and Hoijsink, H. (2005), “Inequality Constrained Analysis of Variance: A Bayesian Approach,” *Psychological Methods*, 10, 477–493. [2,3]
- Lehre, A.-C., Lehre, K. P., Laake, P., and 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. [1]
- Levene, H. (1960), “Robust Tests for Equality of Variances,” in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, eds. I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, Stanford, CA: Stanford University Press, pp. 278–292. [2]
- Lindley, D. V. (1957), “A Statistical Paradox,” *Biometrika*, 44, 187–192. [5]
- Mentz, R. P. (2001), “Likelihood Ratio Test of Equality of Several Variances in the Intraclass Correlation Model,” *Communications in Statistics—Theory and Methods*, 30, 75–91. [2]
- Morgan, W. A. (1939), “A Test for the Significance of the Difference Between the Two Variances in a Sample From a Normal Bivariate Population,” *Biometrika*, 31, 13–19. [2]
- Mulder, J. (2014a), “Bayes Factors for Testing Inequality Constrained Hypotheses: Issues With Prior Specification,” *British Journal of Mathematical and Statistical Psychology*, 67, 153–171. [4,5]
- (2014b), “Prior Adjusted Default Bayes Factors for Testing (In)equality Constrained Hypotheses,” *Computational Statistics and Data Analysis*, 71, 448–463. [2]
- (2016), “Bayes Factors for Testing Order-Constrained Hypotheses on Correlations,” *Journal of Mathematical Psychology*, 72, 104–115. [2]
- Mulder, J., and Fox, J.-P. (2013), “Bayesian Tests on Components of the Compound Symmetry Covariance Matrix,” *Statistics and Computing*, 23, 109–122. [1]
- (2019), “Bayes Factor Testing of Multiple Intraclass Correlations,” *Bayesian Analysis*, 14, 521–552. [1,6,9]
- Mulder, J., Hoijsink, H., and 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. [8,9]
- Mulder, J., Hoijsink, H., and 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. [4]
- Mulder, J., Klugkist, I., van de Schoot, R., Meeus, W. H., Selfhout, M., and Hoijsink, H. (2009), “Bayesian Model Selection of Informative Hypotheses for Repeated Measurements,” *Journal of Mathematical Psychology*, 53, 530–546. [2]
- O’Hagan, A. (1995), “Fractional Bayes Factors for Model Comparison,” *Journal of the Royal Statistical Society, Series B*, 57, 99–138. [2,3]
- Pitman, E. J. G. (1939), “A Note on Normal Correlation,” *Biometrika*, 31, 9–12. [2]
- Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: Wiley. [5]
- (1996), “Multiple Imputation After 18+ Years,” *Journal of the American Statistical Association*, 91, 473–489. [5]
- Schwarz, G. (1978), “Estimating the Dimension of a Model,” *The Annals of Statistics*, 6, 461–464. [2]
- Shoemaker, L. H. (2003), “Fixing the F Test for Equal Variances,” *The American Statistician*, 57, 105–114. [2]
- Tanner, M. A., and Wong, W. H. (1987), “The Calculation of Posterior Distributions by Data Augmentation,” *Journal of the American Statistical Association*, 82, 528–540. [5]
- Vermunt, J. K., and Magidson, J. (2005), “Structural Equation Modeling: Mixture Models,” in *Encyclopedia of Statistics in Behavioral Science*, eds. B. S. Everitt and D. C. Howell, Chichester: Wiley, pp. 1922–1927. [6]
- Wang, M., and Liu, G. (2016), “A Simple Two-Sample Bayesian t -Test for Hypothesis Testing,” *The American Statistician*, 70, 195–201. [2]
- Wilcox, R. R. (1989), “Comparing the Variances of Dependent Groups,” *Psychometrika*, 54, 305–315. [2]
- (1990), “Comparing the Variances of Two Dependent Groups,” *Journal of Educational Statistics*, 15, 237–247. [2]
- (2015), “Comparing the Variances of Two Dependent Variables,” *Journal of Statistical Distributions and Applications*, 2, 1–8. [2]