


Article

New Estimators of the Bayes Factor for Models with High-Dimensional Parameter and/or Latent Variable Spaces

Anna Pajor ^{1,2} 

¹ Department of Mathematics, Cracow University of Economics, ul. Rakowicka 27, 31-510 Kraków, Poland; pajora@uek.krakow.pl or anna.1.pajor@uj.edu.pl

² Department of Financial Mathematics, Jagiellonian University in Kraków, ul. Prof. Stanisława Łojasiewicza 6, 30-348 Kraków, Poland

Abstract: Formal Bayesian comparison of two competing models, based on the posterior odds ratio, amounts to estimation of the Bayes factor, which is equal to the ratio of respective two marginal data density values. In models with a large number of parameters and/or latent variables, they are expressed by high-dimensional integrals, which are often computationally infeasible. Therefore, other methods of evaluation of the Bayes factor are needed. In this paper, a new method of estimation of the Bayes factor is proposed. Simulation examples confirm good performance of the proposed estimators. Finally, these new estimators are used to formally compare different hybrid Multivariate Stochastic Volatility–Multivariate Generalized Autoregressive Conditional Heteroskedasticity (MSV-MGARCH) models which have a large number of latent variables. The empirical results show, among other things, that the validity of reduction of the hybrid MSV-MGARCH model to the MGARCH specification depends on the analyzed data set as well as on prior assumptions about model parameters.

Keywords: Bayesian model comparison; Markov chain Monte Carlo; stochastic volatility



Citation: Pajor, A. New Estimators of the Bayes Factor for Models with High-Dimensional Parameter and/or Latent Variable Spaces. *Entropy* **2021**, *23*, 399. <https://doi.org/10.3390/e23040399>

Academic Editor: Jacinto Martín

Received: 4 March 2021

Accepted: 24 March 2021

Published: 27 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The Bayes factor, defined as a ratio of the marginal likelihoods for the two models being compared, is an important quantity in Bayesian model comparison and hypothesis testing, see, e.g., in [1]. The posterior odds ratio, used for comparing two competing models, is equal to the product of the prior odds and the Bayes' factor. Therefore, one of the challenges for Bayesians is to accurately estimate the factor, especially in models with a large number of parameters and/or latent variables. Most popular methods of calculating the Bayes factor, based on estimation of the marginal likelihoods of each model separately, are very time-consuming (or even computationally infeasible) in high-dimensional cases. Therefore, methods for direct estimation of the Bayes factor, instead of estimating two marginal likelihoods, have been devised. There already exist different approaches, such as the product-space approach, reversible jump MCMC, and path sampling (see, e.g., in [2–5]), which make it possible to estimate the Bayes factor without calculation of the marginal densities of the data, but in many empirical cases (especially in the case of a large number of model parameters) they require extremely complicated and extensively time-consuming operations, with no general guarantees of success.

Note that for nested models with a large number of latent variables such as stochastic volatility models, Jacquier et al. [6] have shown how to estimate the Bayes factor directly using Markov Chain Monte Carlo (MCMC) simulations from the posterior distribution. They pointed out that the Bayes factor can be expressed as the expected value of the ratio of corresponding densities with respect to the posterior distribution of parameters and latent variables. This expected value can be estimated by averaging over the MCMC draws. Thus, using the special structure of stochastic volatility models and exploiting prior assumptions about parameters, they have estimated Bayes factors for comparing the basic stochastic volatility model to the models with fat tails and correlated errors. Unfortunately, the finite

sample properties of the estimator have not been analyzed by the authors. In this paper, we will show that even in very simple models it does not perform well, but it can be improved by trimming the space of parameters and latent variables.

Now, the issue of a direct estimation of the Bayes factor will be described more formally. Let us consider a Bayesian model, in which

1. the space of parameters is denoted by Θ and $p(\theta)$ is a prior density function of the parameters collected in $\theta \in \Theta$,
2. the space of latent variables is denoted by H and $p(h|\theta)$ is a prior density function of the latent variables collected in $h \in H$, and
3. y is a vector of observations.

The marginal data density, $p(y)$, is defined as the integral (calculated over the whole space of parameters and latent variables, $\Theta \times H$) of the conditional data density given the vector of parameters and latent variables, $p(y|\theta, h)$, with respect to the prior distribution of the parameters and latent variables:

$$p(y) = \int_{\Theta \times H} p(y, h, \theta) d\theta dh = \int_{\Theta \times H} p(y|h, \theta) p(h|\theta) p(\theta) d\theta dh. \quad (1)$$

In the majority of models it is impossible to analytically integrate out parameters and latent variables from the joint distribution for y , h , and θ , $p(y, h, \theta)$. It very often results from the number of dimensions of the space of the latent variables and parameters being too large. Furthermore, a correct assessment of the marginal likelihood is computationally challenging (see, e.g., in [7–9], where summaries of various methods can be found). In the presence of a large number of latent variables (e.g., in stochastic volatility models) popular methods of Monte Carlo estimation of $p(y)$ for the observed vector y are computationally extremely intensive, and often they turn out to be infeasible. Fortunately, we can consider the ratio of marginal data densities instead of $p(y)$. In fact, the main criterion of comparison of two competing models M_i and M_j is the posterior odds ratio between the two models:

$$O_{ij} = \frac{p(M_i|y)}{p(M_j|y)} = \frac{p(y|M_i)}{p(y|M_j)} \times \frac{p(M_i)}{p(M_j)}, \quad (2)$$

where $p(M_i|y)$ is the posterior probability of the model M_i , $p(M_i)$ is the prior probability of the model M_i , and $p(y|M_i) = \int_{\Theta_i \times H_i} p(y|h_i, \theta_i, M_i) p(h_i|\theta_i, M_i) p(\theta_i|M_i) d\theta_i dh_i$ is the

marginal data density value (the marginal likelihood) in the model M_i . The ratio $B_{ij} = \frac{p(y|M_i)}{p(y|M_j)}$ is referred to as the Bayes factor in favor of the model M_i against the model M_j ,

in turn, the ratio $\frac{p(M_i)}{p(M_j)}$ is the prior odds ratio. Thus, if the prior odds ratio is equal to 1 (i.e., both models are equally probable a priori, $p(M_i) = p(M_j)$), the posterior odds ratio between the two models is then equal to the Bayes factor. Moreover, if we assume that the set of models $\{M_1, \dots, M_m\}$ is exhaustive, Bayes factors can be used to obtain the posterior model probabilities:

$$p(M_i|y) = \frac{\frac{p(M_i)p(y|M_i)}{p(y|M_k)}}{\frac{\sum_{j=1}^m p(M_j)p(y|M_j)}{p(y|M_k)}} = \frac{p(M_i)B_{ik}}{\sum_{j=1}^m p(M_j)B_{jk}}, \quad (3)$$

for $i, k \in \{1, \dots, m\}$. By choosing $k = i$, we obtain

$$p(M_i|y) = \frac{p(M_i)}{\sum_{j=1}^m p(M_j)B_{ji}}, \quad (4)$$

for $i \in \{1, \dots, m\}$.

Thus, the Bayes factors can be used instead of the marginal data density values. Note that in many situations it is easier to estimate the Bayes factor than the marginal likelihood. In this paper, we show how to estimate the Bayes factor in models with a large number of parameters and/or latent variables, in which calculation of the marginal data density value is numerically impossible. It will be shown that the Bayes factor is equal to the posterior mean, restricted to a certain subset D of the parameter and latent variable space, of the ratio of conditional densities of the corresponding quantities times the reciprocal of the posterior probability of the subset D . This fact leads the researcher to using arithmetic mean estimator of the ratio based on simulation from the posterior distribution restricted to the subset D .

It is well known that the Savage–Dickey density ratio and its generalizations (see in [10–12]) are relatively simple and widely used methods for computing Bayes factors for nested models. The Savage–Dickey method requires an estimate of the value of the posterior distribution at a single point. It is not reliable when this point lies in the tail of the posterior distribution. As has already been mentioned, the Savage–Dickey method can be used only for nested models. Our method can be applied for both nested and non-nested ones.

Note that although the posterior odds principle is fundamental, there are other Bayesian approaches to model comparison or hypothesis testing, e.g., the so-called Lindley type test (see [13]) or the Full Bayesian Significance Test (FBST), introduced by Pereira and Stern [14] as a Bayesian significance test for precise (sharp) hypotheses. A solid theoretical background of the FBST can be found in [15]. Detailed discussions and extensions of the Pereira–Stern test as well as of its applications are presented in, e.g., [16–21]. Our approach to Bayesian model comparison (or comparing hypotheses) is based on posterior probabilities associated with models (or hypotheses). This leads directly to the so-called posterior odds approach. Motivations for the formal Bayesian approach to model selection and for the use of Bayes factors are discussed in [22]. First of all, the Bayes factors and posterior model probabilities are easy to understand. Moreover, this approach is consistent, penalizes model complexity, remains conceptually the same regardless of the number of models (or hypotheses) under consideration, and does not require nested models or regular asymptotics. This approach allows one not only to test hypotheses, but also to compare them—“the process of revising prior probabilities associated with alternative hypotheses does not necessarily involve a decision to reject or accept these hypotheses” (see [1], p. 291). Furthermore, the Bayes factors make it possible to compare models whose prior distributions are different. Finally, the posterior probabilities of models or the Bayes factors can be used in the so-called Bayesian pooling approach (or Bayesian model averaging, see, e.g., in [23]). This paper is organized as follows. In Section 2, our new method is presented. Section 3 contains the simulation study. In Section 4, we present results of comparison of hybrid MSV-MGARCH models. The last section contains conclusions and direction of further research.

2. New Estimators of the Bayes Factor

It is easy to show that the Bayes factor in favor of the model M_i against the model M_j can be expressed as an integral:

$$\frac{p(y|M_i)}{p(y|M_j)} = \int_{\Theta_i \times H_i} \frac{p(y|\theta_i, h_i, M_i)p(h_i|\theta_i, M_i)p(\theta_i|M_i)}{p(y|\theta_j, h_j, M_j)p(h_j|\theta_j, M_j)p(\theta_j|M_j)} p(\theta_j, h_j|y, M_j) d\theta_i dh_i, \quad (5)$$

where $p(y|\theta_i, h_i, M_i)$, $p(h_i|\theta_i, M_i)$, and $p(\theta_i|M_i)$ denote the conditional probability density function of the vector of observations (conditional sampling density), the conditional probability density function of latent variables, and the prior density of parameters in the model M_i , respectively. Note that when the two competing models M_i and M_j have the same parameter vector and the vector of latent variables (i.e., $\theta_i = \theta_j = \theta$, $h_i = h_j = h$, $\Theta_i = \Theta_j = \Theta$, $H_i = H_j = H$), then the Bayes factor can be computed in a relatively

straightforward way which does not require estimation of marginal data density values for each model separately. Of course, it is possible only if draws from one of the posterior distributions are available (see in [24]). We have

$$BF_{ij} = \int_{\Theta \times H} r_{ij}(\theta, h) p(\theta, h|y, M_j) d\theta dh, \tag{6}$$

where

$$r_{ij}(\theta, h) = \frac{p(y|\theta, h, M_i)p(h|\theta, M_i)p(\theta|M_i)}{p(y|\theta, h, M_j)p(h|\theta, M_j)p(\theta|M_j)}.$$

Therefore, given the sample $\{(\theta^{(q)'}, h^{(q)'})\}_{q=1}^k$ from the posterior distribution $p(\theta, h|y, M_j)$, an estimator of BF_{ij} can be expressed as

$$\widehat{BF}_{ij} = \frac{1}{k} \sum_{q=1}^k r_{ij}(\theta^{(q)}, h^{(q)}). \tag{7}$$

As was pointed out in [24], it is crucial that the variability of the ratio $r_{ij}(\theta, h)$ is small under the posterior distribution $p(\theta, h|y, M_j)$; otherwise, estimates of BF_{ij} might be driven by few values of $h^{(q)}$ and $\theta^{(q)}$, and thus an extremely large simulation sample could be required to obtain adequate result. To deal with this problem, we propose a certain modification of the estimator (7). The idea of this modification is based on trimming the posterior sample to a certain subset of parameter and latent variable space, similarly to the idea of the correction of arithmetic mean estimator, proposed in [25]. Let us assume that $D \subseteq \Theta \times H$ and $0 < Pr(D|y, M_i) < \infty$. The equality

$$\frac{p(y|M_i)}{p(y|M_j)} Pr(D|y, M_i) = \int_{\Theta \times H} I_D(\theta, h) \frac{p(y|M_i)}{p(y|M_j)} p(\theta, h|y, M_i) d\theta dh, \tag{8}$$

implies that

$$\frac{p(y|M_i)}{p(y|M_j)} Pr(D|y, M_i) = \int_{\Theta \times H} I_D(\theta, h) \frac{p(y, \theta, h|M_i)}{p(y|M_j)} d\theta dh.$$

We can see immediately that

$$\frac{p(y|M_i)}{p(y|M_j)} Pr(D|y, M_i) = \int_{\Theta \times H} I_D(\theta, h) \frac{p(y|\theta, h, M_i)p(\theta, h|M_i)}{p(y|\theta, h, M_j)p(\theta, h|M_j)} p(\theta, h|y, M_j) d\theta dh,$$

thus

$$\frac{p(y|M_i)}{p(y|M_j)} = \frac{1}{Pr(D|y, M_i)} \int_{\Theta \times H} I_D(\theta, h)r_{ij}(\theta, h) p(\theta, h|y, M_j) d\theta dh. \tag{9}$$

Equality (9) means that the Bayes factor can be expressed as a product of the reciprocal of the posterior probability of the subset D in the model M_i , $Pr(D|y, M_i)$, and of the expected value of the indicator function of subset D times the ratio $r_{ij}(\theta, h)$, $I_D(\theta, h)r_{ij}(\theta, h)$. This expected value is calculated with respect to the posterior distribution of the model parameters and latent variables in the model M_j . Therefore, given the sample $\{(h^{(q)'}, \theta^{(q)'})\}_{q=1}^k$ from the posterior distribution $p(\theta, h|y, M_j)$, an estimator of BF_{ij} can be expressed as

$$\widehat{BF}_{ij,D} = \frac{1}{\widehat{Pr}(D|y, M_i)} \left[\frac{1}{k} \sum_{q=1}^k I_D(\theta^{(q)}, h^{(q)})r_{ij}(\theta^{(q)}, h^{(q)}) \right], \tag{10}$$

where $\hat{Pr}(D|y, M_i)$ is an assessment of the posterior probability of the subset D in the model M_i . Unfortunately, this assessment requires also sampling from the posterior distribution in the model M_i .

Note that equality (9) can obviously be written as

$$\frac{p(y|M_i)}{p(y|M_j)} = \frac{\Pr(D|y, M_j)}{\Pr(D|y, M_i)} \int_D r_{ij}(\theta, h) p(\theta, h|D, y, M_j) d\theta dh, \tag{11}$$

provided that $0 < \Pr(D|y, M_j) < \infty$. This equality suggests that the Bayes factor can be estimated by the product of the ratio of posterior probabilities of the subset D and the sample arithmetic mean of the ratio $r_{ij}(\theta, h)$:

$$\widehat{BF}_{ij,D} = \frac{\hat{Pr}(D|y, M_j)}{\hat{Pr}(D|y, M_i)} \left[\frac{1}{k} \sum_{q=1}^k r_{ij}(\theta^{(q)}, h^{(q)}) \right], \tag{12}$$

based on $\{(h^{(q)}, \theta^{(q)})\}_{q=1}^k$ drawn from the posterior distribution given by $p(\theta, h|y, M_j)$, truncated to the subset D , that is, $p(\theta, h|D, y, M_j)$.

As a by-product of our considerations, we have just shown that if a Monte Carlo sampler only visits a subset of the support of the posterior distribution, the correction of the sample arithmetic mean of the ratio $r_{ij}(\theta, h)$ is needed.

Now, we will extend our analysis to models which contain two groups of parameters: the parameters common to both models and parameters specific to only one of them. Moreover, additional assumptions pertaining to conditional probability density functions will be discussed.

Let us assume that

- (i) $\theta_j = \theta_U = (\theta'_A, \theta'_R)' \in \Theta_U = \Theta_A \times \Theta_R$ denotes all the parameters of the model M_j (since called M_U), while $\theta_i = \theta_R$ contains the parameters common to both models: M_U and M_i (since called M_R), and the vector θ_A denotes specific parameters of M_U ;
- (ii) $p(\theta_U|M_U) = p(\theta_A, \theta_R|M_U) = p(\theta_A|M_U)p(\theta_R|M_U)$, i.e., the random vectors θ_R and θ_A are a priori independent;
- (iii) $\int_{\Theta_A} p(\theta_A|M_U) d\theta_A = 1$, i.e., the prior distribution for the vector of parameters θ_A is proper; and
- (iv) $h_i = h_j = h; H_i = H_j = H$, i.e., competing models have the same vector of latent variables.

2.1. Different Prior Distributions of Latent Variables in Both Models

In this section, we additionally assume that

- (v) $p(y|\theta_U, h_U, M_U) = p(y|\theta_R, h_R, M_R)$, i.e., competing models have the same conditional sampling density and
- (vi) $p(\theta_R|M_U) = p(\theta_R|M_R)$, i.e., in both models, the common parameters have the same prior distribution.

Under above assumptions the Bayes factor in favor of the model M_R versus the model M_U is given by the following integral:

$$\frac{p(y|M_R)}{p(y|M_U)} = \int_{\Theta_A \times \Theta_R \times H} \frac{p(h|\theta_R, M_R)}{p(h|\theta_A, \theta_R, M_U)} p(\theta_A, \theta_R, h|y, M_U) d\theta_A d\theta_R dh. \tag{13}$$

Thus, the estimator of the Bayes factor takes the form

$$\widehat{BF}_{R,U,h} = \frac{1}{k} \sum_{q=1}^k \frac{p(h^{(q)}|\theta_R^{(q)}, M_R)}{p(h^{(q)}|\theta_A^{(q)}, \theta_R^{(q)}, M_U)}, \tag{14}$$

where $\{(\theta_A^{(q)'}, \theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k$ are drawn from the posterior distribution of parameters and latent variables in the model M_U , i.e., from the distribution given by $p(\theta_A, \theta_R, h|y, M_U)$.

On the other hand, it is easy to show that the Bayes factor in favor of the model M_U against the model M_R is as follows:

$$\frac{p(y|M_U)}{p(y|M_R)} = \int_{\Theta_A \times \Theta_R \times H} \frac{p(h|\theta_A, \theta_R, M_U)}{p(h|\theta_R, M_R)} p(\theta_R, h|y, M_R) p(\theta_A|M_U) d\theta_A d\theta_R dh, \tag{15}$$

and consequently

$$\widehat{BF}_{U,R,h} = \frac{1}{k} \sum_{q=1}^k \frac{p(h^{(q)}|\theta_A^{(q)}, \theta_R^{(q)}, M_U)}{p(h^{(q)}|\theta_R^{(q)}, M_R)}, \tag{16}$$

where $\{(\theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k$ and $\{\theta_A^{(q)}\}_{q=1}^k$ are drawn from the posterior distribution $p(\theta_R, h|y, M_R)$ and from the prior distribution, $p(\theta_A|M_U)$, respectively. However, if the dimension of the vector $(\theta', h)'$ is high, then the estimators $\widehat{BF}_{U,R,h}$ and $\widehat{BF}_{R,U,h}$ tend to suffer from the so-called “simulation pseudo-bias”, similarly to the harmonic mean estimator (see in [26]). This “simulation pseudo-bias” of the estimators will be illustrated on the basis of simulation studies in the next section. To deal with the problem of “pseudo-bias”, we propose drawing from the posterior and prior distributions restricted to a subset of the space of parameters and latent variables with non-zero posterior probability, and correcting the arithmetic mean of the ratio by the posterior probability of the subset visited by the sampler.

Let us assume that $D_R \subseteq \Theta_R \times H$ and $D_A \subseteq \Theta_A$, $0 < Pr(D_A \times D_R|M_U) < \infty$, $0 < Pr(D_A \times D_R|y, M_U) < \infty$.

Starting from the identity

$$\begin{aligned} \frac{p(y|M_R)}{p(y|M_U)} Pr(D_R|y, M_R) Pr(D_A|M_U) &= \\ &= \int_{D_A \times D_R} \frac{p(y|M_R)}{p(y|M_U)} p(\theta_R, h|y, M_R) p(\theta_A|M_U) d\theta_A d\theta_R dh, \end{aligned} \tag{17}$$

we obtain

$$\begin{aligned} \frac{p(y|M_R)}{p(y|M_U)} Pr(D_R|y, M_R) Pr(D_A|M_U) &= \\ &= \int_{\Theta_A \times \Theta_R \times H} I_{D_A \times D_R}(\theta_A, \theta_R, h) \frac{p(\theta_R, h, y|M_R) p(\theta_A|M_U)}{p(y|M_U) p(\theta_A, \theta_R, h|y, M_U)} p(\theta_A, \theta_R, h|y, M_U) d\theta_A d\theta_R dh = \\ &= \int_{\Theta_A \times \Theta_R \times H} I_{D_A \times D_R}(\theta_A, \theta_R, h) \frac{p(\theta_R, h, y|M_R) p(\theta_A|M_U)}{p(\theta_A, \theta_R, h, y|M_U)} p(\theta_A, \theta_R, h|y, M_U) d\theta_A d\theta_R dh. \end{aligned}$$

On the basis of the fact that

$$\frac{p(\theta_R, h, y|M_R) p(\theta_A|M_U)}{p(\theta_A, \theta_R, h, y|M_U)} = \frac{p(y|\theta_R, h, M_R) p(h|\theta_R, M_R) p(\theta_R|M_R) p(\theta_A|M_U)}{p(y|\theta_A, \theta_R, h, M_U) p(h|\theta_A, \theta_R, M_U) p(\theta_A, \theta_R|M_U)},$$

and under assumptions (i)–(vi), we have

$$\frac{p(\theta_R, h, y|M_R) p(\theta_A|M_U)}{p(\theta_A, \theta_R, h, y|M_U)} = \frac{p(h|\theta_R, M_R)}{p(h|\theta_A, \theta_R, M_U)}.$$

Therefore,

$$\frac{p(y|M_R)}{p(y|M_U)} = \frac{1}{\Pr(D_R|y, M_R) \Pr(D_A|M_U)} \times \int_{\Theta_A \times \Theta_R \times H} I_{D_A \times D_R}(\theta_A, \theta_R, h) \frac{p(h|\theta_R, M_R)}{p(h|\theta_A, \theta_R, M_U)} p(\theta_A, \theta_R, h|y, M_U) d\theta_A d\theta_R dh. \tag{18}$$

Identity (18) naturally leads to the following estimator of the Bayes factor:

$$\widehat{BF}_{R,U,D,h} = \frac{1}{\hat{Pr}(D_R|y, M_R) \hat{Pr}(D_A|M_U)} \times \sum_{q=1}^k \frac{p(h^{(q)}|\theta_R^{(q)}, M_R)}{p(h^{(q)}|\theta_A^{(q)}, \theta_R^{(q)}, M_U)} I_{D_A \times D_R}(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}), \tag{19}$$

where $\{(\theta_A^{(q)'}, \theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k$ are drawn from $p(\theta_A, \theta_R, h|y, M_U)$, or equivalently

$$\widehat{BF}_{R,U,D,h} = \frac{\hat{Pr}(D_A \times D_R|y, M_U)}{\hat{Pr}(D_R|y, M_R) \hat{Pr}(D_A|M_U)} \times \frac{1}{k} \sum_{q=1}^k \frac{p(h_{D_A \times D_R}^{(q)}|\theta_{R, D_A \times D_R}^{(q)}, M_R)}{p(h_{D_A \times D_R}^{(q)}|\theta_{A, D_A \times D_R}^{(q)}, \theta_{R, D_A \times D_R}^{(q)}, M_U)}, \tag{20}$$

where $\{(\theta_{A, D_A \times D_R}^{(q)'}, \theta_{R, D_A \times D_R}^{(q)'}, h_{D_A \times D_R}^{(q)'})\}_{q=1}^k$ are drawn from $p(\theta_A, \theta_R, h|y, M_U)$ truncated to the subset $D_A \times D_R$. Because the posterior simulation support is the subset $D_A \times D_R$ of the parameter and latent variable space, most of the $\{(\theta_{A, D_A \times D_R}^{(q)'}, \theta_{R, D_A \times D_R}^{(q)'}, h_{D_A \times D_R}^{(q)'})\}_{q=1}^k$ have “similar” values of the ratio of the conditional densities of latent variables, so that, having probabilities $\hat{Pr}(D_A \times D_R|y, M_U)$, $\hat{Pr}(D_R|y, M_R)$, $\hat{Pr}(D_A|M_U)$, the simulation process will be numerically more efficient than in the case of unrestricted space of parameters and latent variables. The estimates will not be dominated by few values of the ratio. Therefore, a smaller simulation sample will be required to obtain adequate precision in $\widehat{BF}_{R,U,D,h}$.

Now suppose that $D \subseteq \Theta_A \times \Theta_R \times H$ and $0 < Pr(D|M_U) < \infty, 0 < Pr(D|y, M_U) < \infty$. This time we start with the identity

$$\frac{p(y|M_U)}{p(y|M_R)} \Pr(D|y, M_U) = \int_D \frac{p(y|M_U)}{p(y|M_R)} p(\theta_A, \theta_R, h|y, M_U) d\theta_A d\theta_R dh, \tag{21}$$

and we obtain

$$\begin{aligned} & \frac{p(y|M_U)}{p(y|M_R)} \Pr(D|y, M_U) = \\ &= \int_{\Theta_A \times \Theta_R \times H} I_D(\theta_A, \theta_R, h) \frac{p(\theta_A, \theta_R, h, y|M_U)}{p(y|M_R) p(\theta_R, h|y, M_R)} p(\theta_R, h|y, M_R) d\theta_A d\theta_R dh = \\ &= \int_{\Theta_A \times \Theta_R \times H} I_D(\theta_A, \theta_R, h) \frac{p(\theta_A, \theta_R, h, y|M_U)}{p(\theta_R, h, y|M_R)} p(\theta_R, h|y, M_R) d\theta_A d\theta_R dh. \end{aligned}$$

In this case,

$$\frac{p(\theta_A, \theta_R, h, y|M_U)}{p(\theta_R, h, y|M_R)} = \frac{p(y|\theta_A, \theta_R, h, M_U) p(h|\theta_A, \theta_R, M_U) p(\theta_A, \theta_R|M_U)}{p(y|\theta_R, h, M_R) p(h|\theta_R, M_R) p(\theta_R|M_R)},$$

and under assumptions (i)–(vi)

$$\frac{p(\theta_A, \theta_R, h, y | M_U)}{p(\theta_R, h, y | M_R)} = \frac{p(h | \theta_A, \theta_R, M_U) p(\theta_A | M_U)}{p(h | \theta_R, M_R)}.$$

Thus,

$$\frac{p(y | M_U)}{p(y | M_R)} = \frac{1}{\Pr(D | y, M_U)} \times \int_{\Theta_A \times \Theta_R \times H} I_D(\theta_A, \theta_R, h) \frac{p(h | \theta_A, \theta_R, M_U)}{p(h | \theta_R, M_R)} p(\theta_R, h | y, M_R) p(\theta_A | M_U) d\theta_A d\theta_R dh. \tag{22}$$

Given a sample $\{(\theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k, \{\theta_A^{(q)}\}_{q=1}^k$ from the distributions $p(\theta_R, h | y, M_R)$ and $p(\theta_A | M_U)$, respectively, then, as results from (22), an estimator of the Bayes factor for the model M_U against the model M_R can be

$$\widehat{BF}_{U,R,D,h} = \frac{1}{\widehat{\Pr}(D | y, M_U)} \frac{1}{k} \sum_{q=1}^k \frac{p(h^{(q)} | \theta_A^{(q)}, \theta_R^{(q)}, M_U)}{p(h^{(q)} | \theta_R^{(q)}, M_R)} I_D(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}). \tag{23}$$

Additionally, if $D = D_A \times D_R \subseteq \Theta_A \times \Theta_R \times H$, then

$$\widehat{BF}_{U,R,D,h} = \frac{\widehat{\Pr}(D_R | y, M_R) \widehat{\Pr}(D_A | M_U)}{\widehat{\Pr}(D | y, M_U)} \frac{1}{k} \sum_{q=1}^k \frac{p(h_D^{(q)} | \theta_{A,D}^{(q)}, \theta_{R,D}^{(q)}, M_U)}{p(h_D^{(q)} | \theta_{R,D}^{(q)}, M_R)}, \tag{24}$$

where $\{(\theta_{R,D}^{(q)'}, h_D^{(q)'})\}_{q=1}^k, \{\theta_{A,D}^{(q)}\}_{q=1}^k$ are drawn from the distributions determined by $p(\theta_R, h | y, M_R)$ and $p(\theta_A | M_U)$, respectively, restricted to the subset D .

2.2. Different Prior Distributions for Common Parameters in Both Models

Now, let us assume that

- (v) $p(y | \theta_U, h_U, M_U) = p(y | \theta_R, h_R, M_R)$, i.e., competing models have the same sampling density and
- (vii) $p(h | \theta_A, \theta_R, M_U) = p(h | \theta_R, M_R)$, i.e., in both models, the latent variables have the same prior distribution, instead of (vi) $p(\theta_R | M_U) = p(\theta_R | M_R)$.

Then, it is easy to show that in the first case

$$\frac{p(y | M_R)}{p(y | M_U)} = \frac{1}{\Pr(D_R | y, M_R) \Pr(D_A | M_U)} \times \int_{\Theta_A \times \Theta_R \times H} I_{D_A \times D_R}(\theta_A, \theta_R, h) \frac{p(\theta_R | M_R)}{p(\theta_R | M_U)} p(\theta_A, \theta_R, h | y, M_U) d\theta_A d\theta_R dh, \tag{25}$$

and in the second case

$$\frac{p(y | M_U)}{p(y | M_R)} = \frac{1}{\Pr(D | y, M_U)} \times \int_{\Theta_A \times \Theta_R \times H} I_D(\theta_A, \theta_R, h) \frac{p(\theta_R | M_U)}{p(\theta_R | M_R)} p(\theta_R, h | y, M_R) p(\theta_A | M_U) d\theta_A d\theta_R dh. \tag{26}$$

Basing on identities (25) and (26), the following estimators of the Bayes factors can be formulated:

$$\widehat{BF}_{R,U,D,\theta} = \frac{1}{\widehat{\Pr}(D_R | y, M_R) \widehat{\Pr}(D_A | M_U)} \frac{1}{k} \sum_{q=1}^k \frac{p(\theta_R^{(q)} | M_R)}{p(\theta_R^{(q)} | M_U)} I_{D_A \times D_R}(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}), \tag{27}$$

where $\{(\theta_A^{(q)'}, \theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k$ are drawn from $p(\theta_A, \theta_R, h|y, M_U)$,

$$\hat{B}F_{U,R,D,\theta} = \frac{1}{\hat{Pr}(D|y, M_U)} \frac{1}{k} \sum_{q=1}^k \frac{p(\theta_R^{(q)}|M_U)}{p(\theta_R^{(q)}|M_R)} I_D(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}), \quad (28)$$

where $\{(\theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k, \{\theta_A^{(q)}\}_{q=1}^k$ are drawn from $p(\theta_R, h|y, M_R)$ and $p(\theta_A|M_U)$, respectively.

2.3. Different Conditional Sampling Distributions of the Vector of Observations

Now we assume that conditions (i)–(iv), (vi), and (vii) hold. Thus, the conditional distribution of the observable vector can be different, but the prior distributions for vector of latent variables and common parameters are the same in both competing models. Then, it is easy to show that

$$\begin{aligned} \frac{p(y|M_R)}{p(y|M_U)} &= \frac{1}{\Pr(D_R|y, M_R) \Pr(D_A|M_U)} \times \\ &\times \int_{\Theta_A \times \Theta_R \times H} I_{D_A \times D_R}(\theta_A, \theta_R, h) \frac{p(y|\theta_R, h, M_R)}{p(y|\theta_A, h, M_U)} p(\theta_A, \theta_R, h|y, M_U) d\theta_A d\theta_R dh, \end{aligned} \quad (29)$$

and for the reciprocal of this Bayes factor we have

$$\begin{aligned} \frac{p(y|M_U)}{p(y|M_R)} &= \frac{1}{\Pr(D|y, M_U)} \times \\ &\times \int_{\Theta_A \times \Theta_R \times H} I_D(\theta_A, \theta_R, h) \frac{p(y|\theta_R, h, M_U)}{p(y|\theta_R, h, M_R)} p(\theta_R, h|y, M_R) p(\theta_A|M_U) d\theta_A d\theta_R dh. \end{aligned} \quad (30)$$

Similarly to the above, basing on identities (29) and (30), the following estimators of the Bayes factors can be formulated:

$$\begin{aligned} \widehat{B}F_{R,U,D,y} &= \frac{1}{\hat{Pr}(D_R|y, M_R) \hat{Pr}(D_A|M_U)} \times \\ &\times \frac{1}{k} \sum_{q=1}^k \frac{p(y|\theta_R^{(q)}, h^{(q)}, M_R)}{p(y|\theta_A^{(q)}, h^{(q)}, M_U)} I_{D_A \times D_R}(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}), \end{aligned} \quad (31)$$

where $\{(\theta_A^{(q)'}, \theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k$ are drawn from $p(\theta_A, \theta_R, h|y, M_U)$, and

$$\widehat{B}F_{U,R,D,y} = \frac{1}{\hat{Pr}(D|y, M_U)} \frac{1}{k} \sum_{q=1}^k \frac{p(y|\theta_R^{(q)}, h^{(q)}, M_U)}{p(y|\theta_R^{(q)}, h^{(q)}, M_R)} I_D(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}), \quad (32)$$

where $\{(\theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k, \{\theta_A^{(q)}\}_{q=1}^k$ are drawn from $p(\theta_R, h|y, M_R)$ and $p(\theta_A|M_U)$, respectively.

Note that the estimators (19), (22), (27), (28), (31), and (32) are based on the arithmetic mean of the ratio of densities times the indicator function of an arbitrary subset of the space of model parameters and latent variables. Additionally, this arithmetic mean is corrected by the reciprocal of the posterior probability of the subset.

3. Simulation Study

In this section, we present a simple simulation study for models with latent variables in which, after having integrated out latent variables analytically, the true values of the marginal likelihoods can easily be assessed using, e.g., the corrected arithmetic mean estimator (CAME [25]). These assessments will be applied to obtain estimates of Bayes factors used as benchmark values for evaluation of the performance of the new method proposed.

Let us consider the Student t model

$$\begin{aligned} y_t | \mu, v &\sim iid(\mu, 1, v), \quad t = 1, 2, \dots, T, \\ \mu &\sim N(\mu_0, \sigma_0^2), v \sim Exp(\lambda_0) \end{aligned} \quad (33)$$

where $t(\mu, 1, v)$ denotes the Student t distribution with mode μ , precision 1, and v degrees of freedom. $N(\mu_0, \sigma_0^2)$ denotes the Normal distribution with mean μ_0 and variance σ_0^2 , in turn, $Exp(\lambda_0)$ stands for the Exponential distribution with mean λ_0^{-1} and variance λ_0^{-2} . The symbol *iid* stands for independent and identically distributed. Thus, the random variables y_1, \dots, y_T are independent and have the same Student t distribution.

The Student t distribution can be expressed as a scale mixture of Gaussian distributions by introducing a random variable h_t that is inverse-gamma distributed. The model can be written as

$$\begin{aligned} y_t | \omega_t, \mu, v &\sim iiN(\mu, h_t^{-1}), \\ h_t | v &\sim iiG(v/2, v/2), \\ \mu &\sim N(\mu_0, \sigma_0^2), v \sim Exp(\lambda_0), \end{aligned} \quad (34)$$

where $G(v/2, v/2)$ denotes the gamma distribution with shape $v/2$ and rate $v/2$.

To simulate datasets we generated samples of size $T = 500, 1000, 2000$, data points from model (33) with $\mu = 0, 0.25, 0.5$, $v = 8$, and $\mu_0 = 0$; $\sigma_0^2 = 1$; $\lambda_0 = 0.1$.

In turn, to simulate from the posterior distribution the Gibbs algorithm was used and run for 20,000 iterations. The conditional posterior distribution for μ is Normal, i.e.,

$$\mu | h, v, y \sim N_K(\mu_1, \sigma_1^2), \quad (35)$$

where $\sigma_1^2 = (\sum_{t=1}^T h_t + \sigma_0^{-2})^{-1}$, $\mu_1 = \sigma_1^2 (\sum_{t=1}^T y_t h_t + \mu_0 \sigma_0^{-2})$, while the conditional posterior distribution for h_t is Gamma:

$$h_t | \mu, v, y \sim G((v+1)/2, [(y_t - \mu)^2 + v]/2), \quad t = 1, \dots, T. \quad (36)$$

An easy computation shows that the conditional posterior distribution of v is nonstandard:

$$v | h, \mu, y \sim p(v | h, y, \mu) \propto \left(\frac{v}{2}\right)^{\frac{Tv}{2}} \Gamma\left(\frac{v}{2}\right)^{-T} e^{-v\kappa/2}, \quad (37)$$

where $\kappa = \sum_{t=1}^T h_t - \sum_{t=1}^T \ln h_t + 2\lambda_0$.

However, reliable numerical methods for generating from this distribution do exist. We use one of them, the rejection sampling proposed by [27].

In the model under consideration, $\theta_U = (\mu, v)'$, $\theta_R = v$, and $h = (h_1, \dots, h_T)'$; thus, the number of latent variables is equal to the number of observations: 500, 1000, and 2000, respectively. The direct computation of the marginal likelihood for model (34) is intensive and unstable due to the presence of latent variables. However, by integrating out the latent variables from the joint distribution of parameters, latent variables, and data, $p(y | h, \mu, v) p(h, \mu, v)$, the conditional density of the data given parameters only can be written in the following form:

$$p(y | \mu, v) = \prod_{t=1}^T \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right) \sqrt{v\pi}} \left(1 + \frac{(y_t - \mu)^2}{v}\right)^{-\frac{v+1}{2}}. \quad (38)$$

Unfortunately, the marginal density of the data (i.e., $p(y)$) cannot be presented in closed form. However, due to the lack of latent variables as well as thanks to the small number of parameters, the marginal data density value can easily and precisely be evalu-

ated by using the corrected arithmetic mean estimator (CAME [25]). Estimates obtained by the use of the CAME are treated as benchmark values.

In order to show performance of our new estimators of the Bayes factor, we assume that the subset D is an intersection of the space of parameters and latent variables, obtained from the condition that the ratio of conditional density functions $p(y|\theta_U, h, M_U)$ and $p(y|\theta_R, h, M_R)$ is between the lowest and the highest values of the ratio evaluated on the basis of pseudo-random sample $\{\theta^{(q)}, h^{(q)}\}_{q=1}^k$ in both models, and the hyper-cuboid limited by the range of the sampler output, i.e.,

$$D = (A \times B) \cap C,$$

where

$$A = \otimes_i [\max_{m \in \{M_U, M_R\}} \min_{q_m} \{\theta_i^{(q_m)}\}, \min_{m \in \{M_U, M_R\}} \max_{q_m} \{\theta_i^{(q_m)}\}],$$

$$B = \otimes_t [\max_{m \in \{M_U, M_R\}} \min_{q_m} \{h_t^{(q_m)}\}, \min_{m \in \{M_U, M_R\}} \max_{q_m} \{h_t^{(q_m)}\}],$$

$$C = \{(\theta'_U, h') : L \leq \frac{p(y|\theta_U, h, M_U)}{p(y|\theta_R, h, M_R)} \leq Q\},$$

$$L = \max_{m \in \{M_U, M_R\}} \min_{q_m} \left\{ \frac{p(y|\theta_U^{(q_m)}, h^{(q_m)}, M_U)}{p(y|\theta_R^{(q_m)}, h^{(q_m)}, M_R)} \right\},$$

$$Q = \min_{m \in \{M_U, M_R\}} \max_{q_m} \left\{ \frac{p(y|\theta_U^{(q_m)}, h^{(q_m)}, M_U)}{p(y|\theta_R^{(q_m)}, h^{(q_m)}, M_R)} \right\}.$$

In the restricted models (M_R), it is assumed that $\mu = 0$, whereas in the unrestricted model (M_U) $\mu \neq 0$. The ratio of density functions of the conditional distributions of the observable vector is as follows:

$$\frac{p(y|\theta_U, h, M_U)}{p(y|\theta_R, h, M_R)} = e^{-\frac{1}{2} \sum_{t=1}^T [(y_t - \mu)^2 - y_t^2] h_t} = e^{-\frac{1}{2} \mu^2 \sum_{t=1}^T h_t + \mu \sum_{t=1}^T y_t h_t}. \tag{39}$$

In Tables 1–3, we present results obtained by using newly proposed estimators of Bayes factors (i.e., the corrected arithmetic means of the ratio of the densities of the conditional distributions of the observable vector, $\widehat{BF}_{U,R,D,y}$ and $\widehat{BF}_{R,U,D,y}$) and uncorrected arithmetic means of the ratios $\widehat{BF}_{U,R,y}$ and $\widehat{BF}_{R,U,y}$. Tables 1–3 present means, standard deviations, root mean square errors and average errors (relative to the CAM estimates) of the decimal logarithm of estimates obtained in models under consideration. As mentioned earlier, closed-form expressions for the marginal likelihoods do not exist. Therefore, the root mean square errors and average errors are determined relative to the CAM estimates obtained for each marginal likelihood separately.

To estimate $\Pr(D|y, M_U)$, $\Pr(D_R|y, M_R)$, and $\Pr(D_A|M_U)$, we use simulation from appropriate posterior or prior distributions, e.g.,

$$\hat{Pr}(D|y, M_U) = \frac{1}{k} \sum_{q=1}^k I_D(\theta_A^{(q)}, \theta_R^{(q)}, h^{(q)}), \tag{40}$$

where $\{(\theta_A^{(q)'}, \theta_R^{(q)'}, h^{(q)'})\}_{q=1}^k$ are drawn from the posterior distribution $p(\theta_A, \theta_R, h|y, M_U)$. The remaining probabilities are calculated in a similar manner. Furthermore, we consider uncorrected arithmetic means of the ratios $\widehat{BF}_{U,R,y}$ and $\widehat{BF}_{R,U,y}$ to investigate sampling properties of these estimators.

Figure 1 indicates that even in such simple models performance of the uncorrected estimators are not satisfactory. The estimator $\widehat{BF}_{R,U,y}$ is downwardly “pseudo-biased” (respective average errors are negative). On the other hand, our simulation study demon-

strates that the performance of the estimators $\widehat{BF}_{U,R,y}$ and $\widehat{BF}_{R,U,y}$ can be improved by trimming the posterior simulation support to a given subset of the space of latent variables and parameters, and next making the correction of the arithmetic mean of the ratios using the posterior probabilities of the subset. As can be seen from Tables 1–3 and Figures 1–3, corrected estimators of the Bayes factors perform very well in comparison to uncorrected ones. The estimators $\widehat{BF}_{U,R,D,y}$ and $\widehat{BF}_{R,U,D,y}$ produce estimates which are very close to those obtained on the basis of the CAME, while the estimators $\widehat{BF}_{U,R,y}$ and $\widehat{BF}_{R,U,y}$, based on uncorrected arithmetic mean, provide biased estimates.

Table 1. Mean (M), standard deviation (SD), average error (AE; corrected arithmetic mean estimator (CAME)—estimated), and root mean square error (RMSE) in the Student t model. The number of observations and of latent variables is equal to $T = 500$.

| Estimator and Its Sample Characteristics | for $\mu = 0$ | | for $\mu = 1/2$ | |
|--|-----------------|-----------------|-----------------|-----------------|
| | $\log BF_{U,R}$ | $\log BF_{R,U}$ | $\log BF_{U,R}$ | $\log BF_{R,U}$ |
| CAME | −1.256 | 1.256 | 21.191 | −21.191 |
| $\log \widehat{BF}_{U,R}$: M | −1.254 | − | 19.826 | − |
| SD | 0.012 | − | 0.748 | − |
| AE | −0.001 | − | −1.340 | − |
| RMSE | 0.011 | − | 1.631 | − |
| $\log \widehat{BF}_{U,R,D}$: M | −1.255 | − | 21.02 | − |
| SD | 0.013 | − | 0.357 | − |
| AE | −0.003 | − | −0.209 | − |
| RMSE | 0.013 | − | 0.589 | − |
| $\log \widehat{BF}_{R,U}$: M | − | 0.498 | − | −22.981 |
| SD | − | 0.165 | − | 0.505 |
| AE | − | −0.776 | − | −1.870 |
| RMSE | − | 0.783 | − | 1.935 |
| $\log \widehat{BF}_{R,U,D}$: M | − | 1.249 | − | −21.238 |
| SD | − | 0.032 | − | 0.284 |
| AE | − | −0.010 | − | −0.100 |
| RMSE | − | 0.036 | − | 0.246 |

Notes: Results obtained for $T = 500$ observations from a Student t distribution with mean μ , precision equal to 1 and 8 degrees of freedom. Estimations of the Bayes factors were repeated 100 times. The Bayes factors were estimated with Monte Carlo sampler based on 20,000 iterations. The $\log BF_{R,U}$ denotes the decimal logarithm of the Bayes factor in favor of model M_R against model M_U . The $\log BF_{U,R}$ denotes the decimal logarithm of the Bayes factor in favor of model M_U against model M_R . $\widehat{BF}_{U,R}$ denotes the estimator of the Bayes factor calculated using simulation from untruncated distributions, whereas $\widehat{BF}_{U,R,D}$ denotes the estimator of the Bayes factor based on simulation from truncated distributions to the set D (the subscript y is omitted for convenience).

Table 2. Mean (M), standard deviation (SD), average error (AE; CAME—estimated), and root mean square error (RMSE) in the Student t model. The number of observations and of latent variables is equal to $T = 1000$.

| Estimator and Its Sample Characteristics | for $\mu = 0$ | | for $\mu = 1/2$ | |
|--|-----------------|-----------------|-----------------|-----------------|
| | $\log BF_{U,R}$ | $\log BF_{R,U}$ | $\log BF_{U,R}$ | $\log BF_{R,U}$ |
| CAME | -1.309 | 1.309 | 37.266 | -37.266 |
| $\log \widehat{BF}_{U,R}$: M | -1.305 | - | 34.806 | - |
| SD | 0.016 | - | 0.738 | - |
| AE | 0.004 | - | -2.485 | - |
| RMSE | 0.020 | - | 2.612 | - |
| $\log \widehat{BF}_{U,R,D}$: M | -1.306 | - | 37.093 | - |
| SD | 0.017 | - | 0.423 | - |
| AE | 0.002 | - | -0.151 | - |
| RMSE | 0.018 | - | 0.403 | - |
| $\log \widehat{BF}_{R,U}$: M | - | 0.415 | - | -40.079 |
| SD | - | 0.201 | - | 0.598 |
| AE | - | -0.813 | - | -2.626 |
| RMSE | - | 0.850 | - | 2.753 |
| $\log \widehat{BF}_{R,U,D}$: M | - | 1.304 | - | -37.341 |
| SD | - | 0.032 | - | 0.258 |
| AE | - | 0.0003 | - | -0.137 |
| RMSE | - | 0.028 | - | 0.265 |

Notes: Results obtained for $T = 1000$ observations from a Student t distribution with mean μ , precision equal to 1 and 8 degrees of freedom. Meaning of symbols used the same as in Table 1.

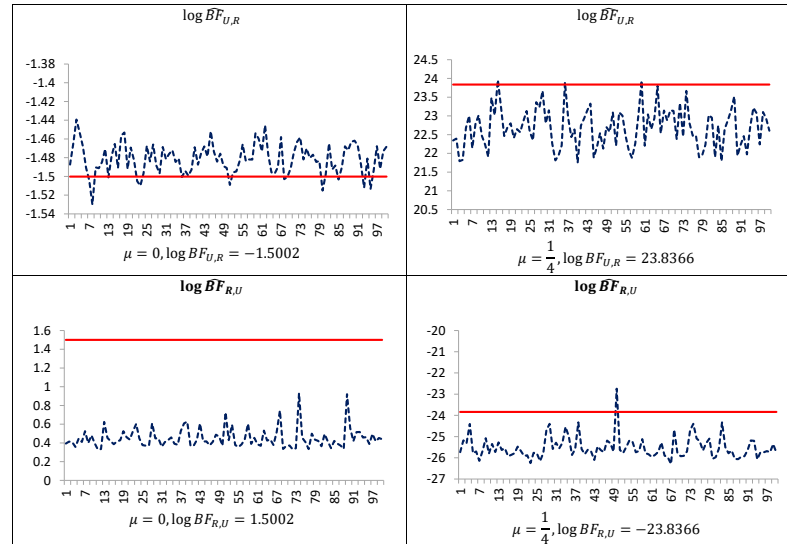


Figure 1. Estimates of the log-Bayes factor in the Student t model, obtained with the use of uncorrected ratios (blue line), and of the corrected arithmetic mean estimator after having integrated out latent variables (red line). Simulation study of the Student t model using 2000 simulated data points ($T = 2000$, which is also equal to the number of latent variables); 20,000 Gibbs sampler iterations for estimation were used. The $\log BF_{R,U}$ denotes the decimal logarithm of the Bayes factor in favor of model M_R against model M_U .

Table 3. Mean (M), standard deviation (SD), average error (AE; CAME—estimated), and root mean square error (RMSE) in the Student t model. The number of observations and of latent variables is equal to $T = 2000$.

| Estimator and Its Sample Characteristics | for $\mu = 0$ | | for $\mu = 1/4$ | |
|--|-----------------|-----------------|-----------------|-----------------|
| | $\log BF_{U,R}$ | $\log BF_{R,U}$ | $\log BF_{U,R}$ | $\log BF_{R,U}$ |
| CAME | −1.500 | 1.500 | 23.837 | −23.837 |
| $\log \widehat{BF}_{U,R}$: M | −1.482 | − | 22.683 | − |
| SD | 0.017 | − | 0.53 | − |
| AE | 0.021 | − | −1.204 | − |
| RMSE | 0.030 | − | 1.321 | − |
| $\log \widehat{BF}_{U,R,D}$: M | −1.480 | − | 23.638 | − |
| SD | 0.016 | − | 0.371 | − |
| AE | 0.024 | − | −0.134 | − |
| RMSE | 0.031 | − | 0.325 | − |
| $\log \widehat{BF}_{R,U}$: M | − | 0.451 | − | −25.521 |
| SD | − | 0.108 | − | 0.517 |
| AE | − | −1.070 | − | −1.715 |
| RMSE | − | 1.072 | − | 1.755 |
| $\log \widehat{BF}_{R,U,D}$: M | − | 1.497 | − | −23.874 |
| SD | − | 0.039 | − | 0.226 |
| AE | − | −0.010 | − | −0.051 |
| RMSE | − | 0.039 | − | 0.283 |

Notes: Results obtained for $T = 2000$ observations from a Student t distribution with mean μ , precision equal to 1 and 8 degrees of freedom. Meaning of symbols used the same as in Table 1.

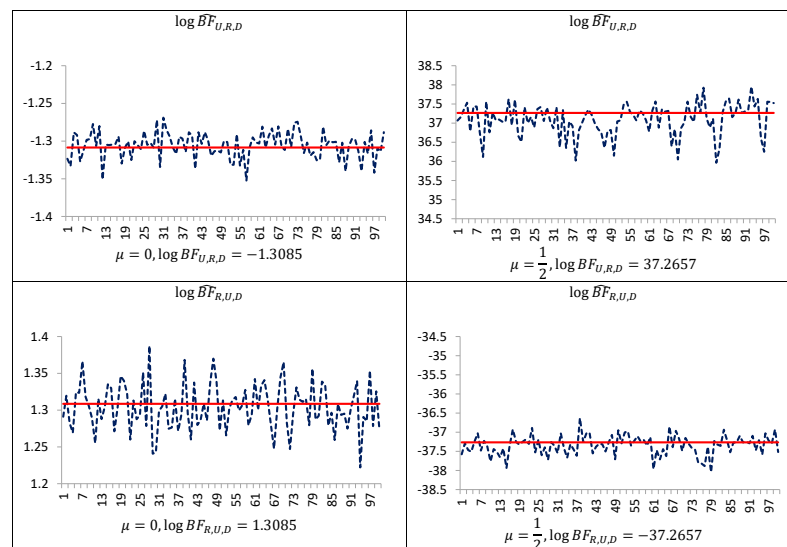


Figure 2. Estimates of the log Bayes factor in the Student t model, obtained with the use of the new estimators (blue line), and of the corrected arithmetic mean estimator after having integrated out latent variables (red line). Simulation study of the Student t model using 1000 simulated data points ($T = 1000$, which is also equal to the number of latent variables); 20,000 Gibbs sampler iterations for estimation were used. The symbol $\widehat{BF}_{R,U,D}$ denotes corrected estimator of the Bayes factor in favor of model M_R against model M_U . The symbol $\widehat{BF}_{U,R,D}$ denotes corrected estimator of the Bayes factor in favor of model M_U against model M_R .

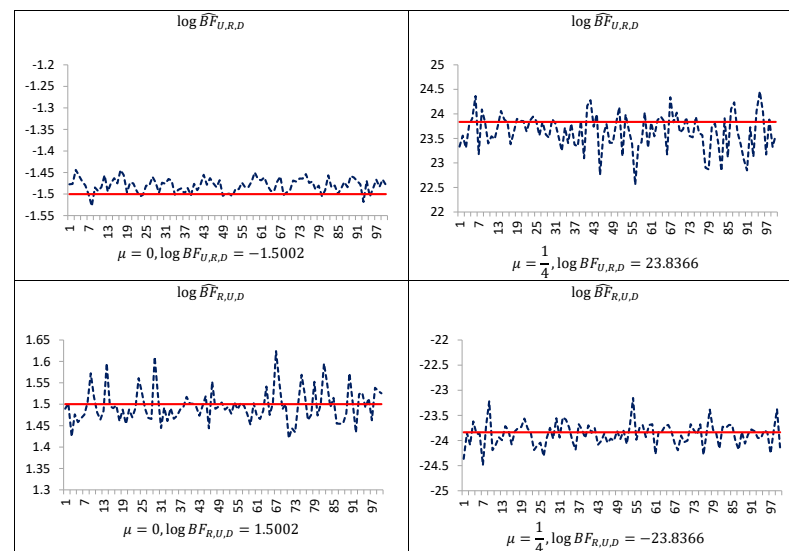


Figure 3. Estimates of the log Bayes factor in the Student t model, obtained with the use of the new estimators (blue line), and of the corrected arithmetic mean estimator after having integrated out latent variables (red line). Simulation study of the Student t model using 1000 simulated data points ($T = 2000$, which is also equal to the number of latent variables); 20,000 Gibbs sampler iterations for estimation were used. The symbol $\widehat{BF}_{R,U,D}$ denotes corrected estimator of the Bayes factor in favor of model M_R against model M_U . The symbol $\widehat{BF}_{U,R,D}$ denotes corrected estimator of the Bayes factor in favor of model M_U against model M_R .

Finally, note that by using both estimators for the ratio of densities and their reciprocals, i.e., $\widehat{BF}_{U,R,D,y}$ and $\widehat{BF}_{R,U,D,y}$, one can check the accuracy of assessments and of adequate (from numerical point of view) selection of the subset D . This is because, roughly speaking, the equality $BF_{U,R} = \frac{1}{BF_{R,U}}$ implies that it is natural to use $\widehat{BF}_{U,R,D,y}$ and $\frac{1}{\widehat{BF}_{R,U,D,y}}$ to estimate $BF_{U,R}$. Different estimates for $BF_{U,R}$ can indicate numerical inadequacy of selection of the subset D and/or too small simulation sample.

4. Empirical Illustration: Formal Bayesian Comparison of Hybrid MSV-MGARCH Models

In this section, the new method will be applied in order to formally compare the hybrid Multivariate Stochastic Volatility–Multivariate Generalized Autoregressive Conditional Heteroskedasticity (MSV-MGARCH) models and thus to show that it can be used in practice. The hybrid MSV-MGARCH models were proposed in [28–30] for modeling financial time series. These hybrid models are characterized by relatively simple model structures that exploit advantages of both model classes: flexibility of the Multivariate Stochastic Volatility (MSV) class, where volatility is modeled by latent stochastic processes, and relative simplicity of the Multivariate GARCH (MGARCH) class. The simplest specification of MSV-MGARCH model (called LN-MSF-SBEKK) is based on one latent process (Multiplicative Stochastic Factor (MSF)) and on the scalar BEKK [31] covariance structure. The LN-MSF-SBEKK structure is obtained by multiplying the SBEKK conditional covariance matrix H_t by a scalar random variable h_t such that $\{\ln h_t\}$ is a Gaussian AR(1) latent process with autoregression parameter ϕ . The hybrid LN-MSF-SBEKK specification has been recognized in the literature [32–35] and proved to be useful in multivariate modeling of financial time series as well as of macroeconomic data [36–41]. The drawback of the LN-MSF-MGARCH process is that it cannot be treated as a direct extension of the MGARCH process with the Student t conditional distribution. When $\phi = 0$, the LN-MSF-SBEKK process reduces itself to the SBEKK process with the conditional distribution being a continuous mixture of multivariate normal distributions with covariance matrices $H_t h_t$ and h_t log-normally distributed. However, the multivariate Student t distribution can be expressed as a scale mixture of normal distributions with the inverted gamma as a mixing

distribution. This fact was exploited in [42,43], where the IG-MSF-SBEKK specification was proposed as a natural hybrid extension of the SBEKK process with the Student t conditional distribution (t-SBEKK). In the new specification, the latent process $\{\ln h_t\}$ remains an autoregressive process of order one, but it is non-Gaussian. For $\phi = 0$ the latent variables h_t (where $t \in \mathbb{Z}$) are independent and have inverted gamma (IG) distribution. Unfortunately, for $\phi \neq 0$ the unconditional distribution of the latent variables h_t is unknown. To summarize, the IG-MSF-SBEKK model is obtained by multiplying the SBEKK conditional covariance matrix H_t by a scalar random variable h_t coming from such latent process (with autoregression parameter ϕ) that, for $\phi = 0$, h_t has an inverted gamma distribution. Thus, $\phi = 0$ leads to the t-SBEKK specification, in which the conditional distribution is represented as a continuous mixture of multivariate normal distributions with covariance matrices $H_t h_t$, where h_t is inverted gamma distributed. If $\phi \neq 0$, the latent variables h_t ($t \in \mathbb{Z}$) are dependent, so (in comparison to the t-SBEKK model) in the IG-MSF-SBEKK model there exists an additional source of dependence.

Let us consider a two-dimensional autoregressive process $r_t = (r_{1,t}, r_{2,t})$ defined by the equation

$$r_t = \delta_0 + r_{t-1}\Delta + \varepsilon_t, \quad t = 1, \dots, T, \quad (41)$$

where δ_0 and Δ are, respectively, 2×1 and 2×2 matrix parameters, and T is the length of the observed time series. The hybrid MSF-MGARCH specification for the disturbance term ε_t is defined by the following equality:

$$\varepsilon_t = \zeta_t H_t^{1/2} h_t^{1/2}, \quad (42)$$

where

$$H_t = (1 - \beta_1 - \beta_2)A + \beta_1(\varepsilon_{t-1}'\varepsilon_{t-1}) + \beta_2 H_{t-1}, \quad (43)$$

$$\ln g_t = \phi \ln g_{t-1} + \ln \gamma_t, \quad (44)$$

$\{\zeta_t\}$ is a Gaussian white noise sequence with mean vector zero and unit covariance matrix; $\{\gamma_t\}$ is a sequence of independent positive random variables; γ_t is inverted gamma distributed with mean $\frac{v}{v-2}$ for $v > 2$, i.e., $\{\gamma_t\} \sim iiIG(v/2, v/2)$, $\zeta_t \perp \gamma_s$ for all $t, s \in \{1, \dots, T\}$, $0 < |\phi| < 1$; β_1 and β_2 are positive scalar parameters such that $\beta_1 + \beta_2 < 1$; and A is a free symmetric positive definite matrix of order 2. Under (41) and (42), the conditional distribution of r_t (given the past of r_t and the current latent variable h_t) is Normal with mean vector $\mu_t = \delta_0 + r_{t-1}\Delta$ and covariance matrix $\Sigma_t = H_t h_t$. For $\phi = 0$ (this value is excluded in the definition of the hybrid models under consideration) $h_t = \gamma_t$, so the distribution of h_t is an inverted gamma. In this case, r_t in (41) is, given its past, an IG mixture of two-variate normal distributions $N(\mu_t, h_t H_t)$, i.e., it has the two-variate Student t distribution.

The assumptions presented so far determine the conditional distribution of the observations and latent variables given the parameters. Thus, it remains to formulate the prior distributions of parameters. We use the same prior distributions as in [42,43]. Six elements of $\delta = (\delta_0 \text{vec}(\Delta))'$ are assumed to be a priori independent of other parameters, with the $N(0, I_6)$ prior. Matrix A has an inverted Wishart prior distribution such that A^{-1} has the Wishart prior distribution with mean I_2 ; the elements of $\beta = (\beta_1, \beta_2)'$ are jointly uniformly distributed over the unit simplex. As regards the initial conditions for H_t , we take $H_0 = h_0 I_2$ and treat $h_0 > 0$ as an additional parameter, a priori exponentially distributed with mean 1; ϕ has the uniform distribution over $(-1, 1)$, and for v we assume the gamma distribution with mean λ_a / λ_v , truncated to $(2, +\infty)$. We assume that $\lambda_v = 0.1$ with two cases: $\lambda_a = 3$ and $\lambda_a = 1$ (note that $\lambda_a = 1$ leads to exponential distribution for v).

Furthermore, we use the same bivariate data sets as those modeled in [30,42,43]. The first data set consists of the official daily exchange rates of the National Bank of Poland (NBP fixing rates) for the US dollar and German mark in the period 1 February 1996–31 December 2001. The length of the modeled time series of their daily growth rates (logarithmic return rates) is 1482. The second data set consists of the daily quotations of the

main index of the Warsaw Stock Exchange (WIG) and the S&P500 index of NYSE—1727 logarithmic returns are modeled from the period 8 January 1999–1 February 2006.

In order to obtain pseudo-random sample from the posterior distribution of parameters and latent variables, we use MCMC simulation techniques described in [42,44] and implemented within the GAUSS programming environment.

Using fully Bayesian methods, we want to answer the question whether the IG-MSF-SBEKK model can be reduced to the t-SBEKK case. Thus, we consider the hypothesis that a scalar parameter $\phi = 0$ (the t-SBEKK model, M_R) and the alternative hypothesis that $\phi \neq 0$ (the IG-MSF-SBEKK model, M_U). For the exchange rate data, the posterior probability that $\phi < 0$ is approximately 0.001 only and $\phi = 0$ is included in the highest posterior density (HPD) interval of probability content as high as 0.996. Thus, Lindley’s procedure leads to the conclusion that the t-SBEKK is inadequate. But for the stock data, the posterior probability that $\phi < 0$ is 0.017 for $\lambda_a = 3$ and 0.054 for $\lambda_a = 1$, and $\phi = 0$ is included in the HPD interval of probability content 0.87 or 0.80, depending on the prior hyperparameter λ_a . In the case of the stock data, Lindley’s testing procedure yields results that the t-SBEKK model cannot be rejected by the data.

Now, our new estimators of the Bayes factors will be used. We start with the assumption that the subset D is an intersection of the subspace of parameters and latent variables, obtained by the condition that the ratio of conditional density functions $p(h|\theta_U, M_U)$ and $p(h|\theta_R, M_R)$ is between the lowest and the highest values of the ratio evaluated at pseudo-random sample $\{\theta^{(q)}, h^{(q)}\}_{q=1}^k$ in both models, and the hyper-cuboid limited by the range of the sampler output, i.e., $D = (A \times B) \cap C$, where

$$A = \otimes_i \left[\max_{m \in \{M_U, M_R\}} \min_{q_m} \{\theta_i^{(q_m)}\}, \min_{m \in \{M_U, M_R\}} \max_{q_m} \{\theta_i^{(q_m)}\} \right],$$

$$B = \otimes_t \left[\max_{m \in \{M_U, M_R\}} \min_{q_m} \{h_t^{(q_m)}\}, \min_{m \in \{M_U, M_R\}} \max_{q_m} \{h_t^{(q_m)}\} \right],$$

$$C = \{(\theta'_U, h') : L \leq \frac{p(h|\theta_U, M_U)}{p(h|\theta_R, M_R)} \leq Q\},$$

$$L = \max_{m \in \{M_U, M_R\}} \min_{q_m} \left\{ \frac{p(h^{(q_m)}|\theta_U^{(q_m)}, M_U)}{p(h^{(q_m)}|\theta_R^{(q_m)}, M_R)} \right\},$$

$$Q = \min_{m \in \{M_U, M_R\}} \max_{q_m} \left\{ \frac{p(h^{(q_m)}|\theta_U^{(q_m)}, M_U)}{p(h^{(q_m)}|\theta_R^{(q_m)}, M_R)} \right\}.$$

The ratio of the densities of the conditional distributions of the vector of latent variables is as follows:

$$\frac{p(h|\theta_U, M_U)}{p(h|\theta_R, M_R)} = \prod_{t=1}^T h_{t-1}^{\frac{v}{2}\phi} e^{\frac{v}{2h_t}(1-h_{t-1}^\phi)}. \tag{45}$$

In Table 4, we present results obtained by using newly proposed estimators of Bayes factors—the corrected arithmetic means of the ratio of the densities of the conditional distributions of latent variables, $\widehat{BF}_{U,R,D}$ and $\widehat{BF}_{R,U,D}$; the subscript h is omitted for convenience. Results obtained on the basis of our new method confirm that in the case of exchange rates, the t-SBEKK model is strongly rejected by the data, whereas it seems that the t-SBEKK specification can be adequate for stock indices. Under equal prior model probabilities, the IG-MSF-SBEKK model for exchange rates is about 14–15 times more probable *a posteriori* than the t-SBEKK model, indicating a strong (but not very strong) evidence against the t-SBEKK model. In turn, for stock indices the decimal logarithm of the Bayes factor of t-SBEKK relative to IG-MSF-SBEKK depends on prior distribution of the number of degrees of freedom, v . The Bayes factor in favor of t-SBEKK is equal to about 1.25 and 0.4 in models with $\lambda_a = 1$ and $\lambda_a = 3$, respectively. Of course, according to scale of Kass and Raftery

(1995) these results indicate evidence for a given model that is negligible: “not worth more than a bare of mention”.

Table 4. Estimates of Bayes factors for the IG-MSF-SBEKK (M_U) and t-SBEKK (M_R) models.

| Estimator | Exchange Rates, $T = 1482$ | | Stock Indices, $T = 1742$ | |
|--------------------------------------|----------------------------|-------------------|---------------------------|-------------------|
| | $\lambda_a = 1$ | $\lambda_a = 3$ | $\lambda_a = 1$ | $\lambda_a = 3$ |
| $\log \widehat{BF}_{U,R,D}$ (NSE) | 1.147 (0.017) | 1.191 (0.024) | -0.110 (0.012) | 0.400 (0.014) |
| $\log \widehat{BF}_{R,U,D}$ (NSE) | -1.134 (0.009) | -1.182 (0.019) | 0.092 (0.011) | -0.388 (0.024) |
| log SD ratio | -1.159 | -1.167 | 0.118 | -0.396 |

Notes: The Bayes factors were estimated with Monte Carlo sampler based on 2,000,000 iterations. The $\log \widehat{BF}_{R,U,D}$ denotes the decimal logarithm of the estimator of the Bayes factor in favor of model M_R against model M_U . NSE denotes the numerical standard error and the log SD ratio stands for the decimal logarithm of the Savage–Dickey density ratio. Scale for the strength of evidence against M_R [45]: $0 < \log BF_{U,R} \leq 1/2$ —negligible; $1/2 < \log BF_{U,R} \leq 1$ —mild; $1 < \log BF_{U,R} \leq 2$ —strong; $2 < \log BF_{U,R}$ —very strong.

For the sake of comparison, in the last row of the Table 4 the estimates of the Savage–Dickey ratio is presented. The marginal posterior density of ϕ is evaluated at $\phi = 0$ on the basis of the histogram of the sampled values of ϕ from the posterior distribution. The main conclusions pertaining to the validity of the reduction of the IG-MSF-SBEKK model to the t-SBEKK one are very similar.

Additionally, it is very important to stress that our method makes it possible to compare IG-MSF-SBEKK models whose prior distributions are different (e.g., in respect of the number of degrees of freedom). As can be seen from Table 5, for exchange rates the IG-MSF-SBEKK and t-SBEKK models with exponential distribution for v are more probable a posteriori than those with the gamma distribution for v with the hyperparameter $\lambda_a = 3$. In contrary, for stock indices the IG-MSF-SBEKK model with the hyperparameter $\lambda_a = 3$ is more probable a posteriori than the same model with $\lambda_a = 1$, but this improvement is negligible.

Table 5. Estimates of Bayes factors for the IG-MSF-SBEKK models with different prior distributions of the number of degrees of freedom.

| Estimator | Exchange Rates, $T = 1482$ | | Stock Indices, $T = 1742$ | |
|---|---------------------------------|-------------------------|---------------------------------|-------------------------|
| | $\phi \neq 0$ (IG-MSF-SBEKK) | $\phi = 0$ (t-SBEKK) | $\phi \neq 0$ (IG-MSF-SBEKK) | $\phi = 0$ (t-SBEKK) |
| $\widehat{BF}_{\lambda_a=1,\lambda_a=3,D}$ | 8.123 | 9.517 | 0.524 | 1.487 |
| $\log \widehat{BF}_{\lambda_a=1,\lambda_a=3,D}$ | 0.910 | 0.979 | -0.281 | 0.172 |
| $\widehat{BF}_{\lambda_a=3,\lambda_a=1,D}$ | 0.123 | 0.105 | 1.850 | 0.622 |
| $\log \widehat{BF}_{\lambda_a=3,\lambda_a=1,D}$ | -0.909 | -0.978 | 0.269 | -0.206 |

Notes: The Bayes factors were estimated with Monte Carlo sampler based on 2,000,000 iterations. The $\log \widehat{BF}_{\lambda_a=1,\lambda_a=3,D}$ denotes the decimal logarithm of the estimator of the Bayes factor in favor of the IG-MSF-SBEKK model with $\lambda_a = 1$ against the IG-MSF-SBEKK model with $\lambda_a = 3$.

5. Discussion

In this paper, a new method of the estimation of the Bayes factor is proposed. The idea of proposed estimators is based on correction of the arithmetic mean estimator of the ratio of conditional distributions by trimming the posterior sample to a certain subset of the space of parameters and latent variables, D , and correcting the arithmetic mean by the posterior probabilities of this subset. The new method makes it possible to compare a finite number of different models with a large number of parameters and/or latent variables in respect to the goodness of fit measured by their posterior probabilities. Our simulation study and empirical illustration show that the method performs well.

In this paper, the question of an adequate selection of the subset D used in the method proposed is not addressed. The choice of the subset which minimizes the variance of the estimator remains an open question. The simulation study and empirical example considered in the paper indicate that the choice of D as an intersection of the parameter space (with additional restrictions) and the hyper-cuboid limited by the range of the posterior sampler outputs leads to acceptable results.

Funding: Research supported by a grant from the National Science Center in Poland under decision no. UMO-2018/31/B/HS4/00730.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are publicly available.

Conflicts of Interest: The author declares no conflicts of interest.

References

- Zellner, A. *An Introduction to Bayesian Inference in Econometrics*; John Wiley and Sons, Inc.: New York, NY, USA, 1971.
- Gelman, A.; Meng, X.L. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Stat. Sci.* **1998**, *13*, 163–185. [[CrossRef](#)]
- Green, P.J. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **1995**, *82*, 711–732. [[CrossRef](#)]
- Han, C.; Carlin, B.P. Markov chain Monte Carlo methods for computing Bayes factors: A comparative review. *J. Am. Stat. Assoc.* **2001**, *96*, 1122–1132. [[CrossRef](#)]
- Sinharay, S.; Stern, H.S. An empirical comparison of methods for computing Bayes factors in generalized linear mixed models. *J. Comput. Graph. Stat.* **2005**, *14*, 415–435. [[CrossRef](#)]
- Jacquier, E.; Polson, N.G.; Rossi, P.E. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *J. Econom.* **2004**, *122*, 185–212. [[CrossRef](#)]
- Ardia, D.; Baştürk, N.; Hoogerheide, L.F.; van Dijk, H.K. A Comparative Study of Monte Carlo Methods for Efficient Evaluation of Marginal Likelihood. *Comput. Stat. Data Anal.* **2012**, *56*, 3398–3414. [[CrossRef](#)]
- Chen, M.H.; Shao, Q.M.; Ibrahim, J.G. *Monte Carlo Methods in Bayesian Computation*; Springer: New York, NY, USA, 2012.
- Llorente, F.; Martino, L.; Delgado, D.; Lopez-Santiago, J. Marginal Likelihood Computation for Model Selection and Hypothesis Testing: An Extensive Review. Available online: <https://arxiv.org/abs/2005.08334> (accessed on 3 March 2021).
- Cameron, E. A generalized Savage-Dickey density ratio. *arXiv* **2013**, arXiv:1311.1292.
- Dickey, J. The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *Ann. Math. Stat.* **1971**, *42*, 204–223. [[CrossRef](#)]
- Verdinelli, I.; Wasserman, L. Computing Bayes factors by using a generalization of the Savage-Dickey density ratio. *J. Am. Stat. Assoc.* **1995**, *90*, 614–618. [[CrossRef](#)]
- Lindley, D.V. *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 2. Inference*; Cambridge University Press: Cambridge, UK, 1965.
- de Bragança Pereira, C.; Stern, J.M. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy* **1999**, *1*, 99–110. [[CrossRef](#)]
- Borges, W.; Stern, J.M. The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic J. IGPL* **2007**, *15*, 401–420. [[CrossRef](#)]
- Diniz, M.A.; Pereira, C.A.B.; Stern, J.M. Cointegration and Unit Root Tests: A Fully Bayesian Approach. *Entropy* **2020**, *22*, 968. [[CrossRef](#)]
- Esteves, L.G.; Izicki, R.; Stern, J.M.; Stern, R.B. The Logical Consistency of Simultaneous Agnostic Hypothesis Tests. *Entropy* **2016**, *18*, 256. [[CrossRef](#)]
- Lauretto, M.S.; Faria, S.R.; Pereira, C.A.B.; Stern, J.M. The Problem of Separate Hypotheses via Mixtures Models. *Am. Inst. Phys. Conf. Proc.* **2007**, *954*, 268–275.
- Madruça, M.R.; Esteves, L.G.; Wechsler, S. On the Bayesianity of Pereira-Stern tests. *Test* **2001**, *10*, 291–299. [[CrossRef](#)]
- Madruça, M.R.; Pereira, C.A.B.; Stern, J.M. Bayesian Evidence Test for Precise Hypotheses. *J. Stat. Plan. Inference* **2003**, *117*, 185–198. [[CrossRef](#)]
- Pereira, C.A.B.; Stern, J.M.; Wechsler, S. Can a significance test be genuinely Bayesian? *Bayesian Anal.* **2008**, *3*, 79–100. [[CrossRef](#)]
- Berger, J.O.; Pericchi, C.R. Objective Bayesian methods for model selection: introduction and comparison (with discussion). In *Model Selection; Lecture Notes—Monograph Series 38*; Lahiri, P., Ed.; Institute of Mathematical Statistics: Hayward, CA, USA, 2001; pp. 135–207.

23. Osiewalski, J.; Steel, M. Una perspectiva bayesiana en selección de modelos. *Cuad. Econ.* **1993**, *55*, 327–351. Available online: <https://home.cyf-kr.edu.pl/eosiewa/ABayesianPerspectiveonModelSelection.pdf> (accessed on 17 March 2021).
24. Chan, J.C.C.; Eisenstat, E. Marginal Likelihood Estimation with the Cross-Entropy Method. *Econom. Rev.* **2015**, *34*, 256–285. [[CrossRef](#)]
25. Pajor, A. Estimating the Marginal Likelihood Using the Arithmetic Mean Identity. *Bayesian Anal.* **2017**, *12*, 261–287. [[CrossRef](#)]
26. Lenk, P. Simulation pseudo-bias correction to the harmonic mean estimator of integrated likelihoods. *J. Comput. Graph. Stat.* **2009**, *18*, 941–960. [[CrossRef](#)]
27. Geweke, J. Priors for macroeconomic time series and their application. In *Institute for Empirical Macroeconomics Discussion Paper 64*; Federal Reserve Bank of Minneapolis: Minneapolis, MN, USA, 1992. Available online: <https://www.minneapolisfed.org/research/dp/dp64.pdf> (accessed on 27 March 2021).
28. Osiewalski, J. New hybrid models of multivariate volatility (a Bayesian perspective). *Przegląd Stat. Stat. Rev.* **2009**, *56*, 15–22.
29. Osiewalski, J.; Pajor, A. Flexibility and parsimony in multivariate financial modelling: A hybrid bivariate DCC–SV model. In *Financial Markets. Principles of Modeling, Forecasting and Decision-Making*; FindEcon Monograph Series No.3; Milo, M., Wdowiński, P., Eds.; Łódź University Press: Łódź, Poland, 2007; pp. 11–26.
30. Osiewalski, J.; Pajor, A. Bayesian analysis for hybrid MSF–SBEKK models of multivariate volatility. *Cent. Eur. J. Econ. Model. Econom.* **2009**, *1*, 179–202.
31. Baba, Y.; Engle, R.; Kraft, D.; Kroner, K. *Multivariate Simultaneous Generalised ARCH*; Department of Economics, University of California at San Diego: San Diego, CA, USA, 1989.
32. Amado, C.; Teräsvirta, T. Modelling volatility by variance decomposition. *J. Econom.* **2013**, *175*, 142–153. [[CrossRef](#)]
33. Amado, C.; Teräsvirta, T. Specification and testing of multiplicative time-varying GARCH models with applications. *Econom. Rev.* **2017**, *36*, 421–446. [[CrossRef](#)]
34. Carriero, A.; Clark, T.; Marcellino, M. Common drifting volatility in large Bayesian vars. *J. Bus. Econ. Stat.* **2016**, *34*, 375–390. [[CrossRef](#)]
35. Teräsvirta, T. Nonlinear models for autoregressive conditional heteroscedasticity. In *Handbook of Volatility Models and Their Applications*; Bauwens, L., Hafner, C., Laurent, S., Eds.; Wiley: New York, NY, USA, 2012; pp. 49–69.
36. Osiewalski, K.; Osiewalski, J. A long-run relationship between daily prices on two markets: The Bayesian VAR(2)-MSF-SBEKK model. *Cent. Eur. J. Econ. Model. Econom.* **2013**, *5*, 65–83.
37. Osiewalski, J.; Osiewalski, K. Hybrid MSV-MGARCH models—General remarks and the GMSF-SBEKK specification. *Cent. Eur. J. Econ. Model. Econom.* **2016**, *8*, 241–271.
38. Osiewalski, J.; Pajor, A. Bayesian Value-at-Risk for a portfolio: Multi- and univariate approaches using MSF-SBEKK models. *Cent. Eur. J. Econ. Model. Econom.* **2010**, *2*, 253–277.
39. Pajor, A.; Osiewalski, J. Bayesian Value-at-Risk and Expected Shortfall for a large portfolio (multi- and univariate approaches). *Acta Phys. Pol. A* **2012**, *121*, 101–109. [[CrossRef](#)]
40. Pajor, A.; Wróblewska, J. VEC-MSF models in Bayesian analysis of short- and long-run relationships. *Stud. Nonlinear Dyn. E* **2017**, *21*, 1–22. [[CrossRef](#)]
41. Wróblewska, J.; Pajor, A. One-period joint forecasts of Polish inflation, unemployment and interest rate using Bayesian VEC-MSF models. *Cent. Eur. J. Econ. Model. Econom.* **2019**, *11*, 23–45.
42. Osiewalski, J.; Pajor, A. A hybrid MSV-MGARCH generalisation of the *t*-MGARCH model. In *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena: Conference Proceedings, No. 1*; Papież, M., Śmiech, S., Eds.; Foundation of the Cracow University of Economics: Cracow, Poland, 2018; pp. 345–354.
43. Osiewalski, J.; Pajor, A. On Sensitivity of Inference in Bayesian MSF-MGARCH Models. *Cent. Eur. J. Econ. Model. Econom.* **2019**, *11*, 173–197.
44. Pajor, A. MCMC Method for the IG-MSF-SBEKK Model. In *Quantitative Methods in the Contemporary Issues of Economics*; Ciałowicz, B., Ed.; Wydawnictwo edu-Libri s.c.: Kraków, Poland, 2020; pp. 77–88. Available online: <http://Quantitative-methods.pdf> (accessed on 3 March 2021).
45. Kass, R.E.; Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [[CrossRef](#)]