Test manuscript No.
(will be inserted by the editor)

Combining Bayesian procedures for testing

Rosangela H. Loschi · Cristiano C. Santos ·
Reinaldo B. Arellano-Valle.

Received: February 19, 2009/ Accepted: date

Abstract Jeffreys and Pereira-Stern Bayesian procedures for testing provide measures of evidence in favour the null hypothesis which can lead to different decisions. We introduce two procedures for testing based on pooling the posterior evidences in favour of the null hypothesis provided by these procedures. We prove that the proposed procedure which has been built using the linear pool of probability is a Bayes test and does not lead to Jeffreys-Lindley paradox. We apply the results for testing precise hypothesis about parameters of some asymmetric family of distributions including the skew-normal one.

Keywords Asymmetric distributions · Bayes risk · Bayes tests · opinion pools · precise hypothesis

Mathematics Subject Classification (2000) 62C10 · 62F15 · 62F03

1 Introduction

Jeffreys [11] introduces a procedure for testing that is based on the Bayes factor. Bayes factor is defined as the ratio of the posterior to the prior odds in favour of the null hypothesis H_0. We decide in favour of H_0, whenever the Bayes factor $BF(H_0, H_1)$ assumes high value. Equivalently, we can make our decision based on the posterior of H_0, $P(H_0|\mathbf{x})$, which is a function of the Bayes factor. This procedure is usually named Jeffreys test.

More recently, Pereira and Stern [19] introduce another measure of evidence in favour of H_0. This measure of evidence is the region over the posterior obtained considering all points of the parametric space for which the posterior values are, at most, as large as the supremum over the subset of the parametric space Θ induced by the

R. H. Loschi and C. C. Santos
Departamento de Estatística, Universidade Federal de Minas Gerias, Brazil
Tel.: +55-31-34095938
Fax: +55-31-34095924
E-mail: loschi@est.ufmg.br

R. B. Arellano-Valle
Departamento de Estadística, Pontificia Universidad Católica de Chile, Chile

null hypothesis. Therefore, Pereira-Stern measure of evidence in favour of the null hypothesis, $Ev(H_0, \mathbf{x})$, is the posterior probability related to the less probable points of Θ. The decision is in favour of H_0, whenever $Ev(H_0, \mathbf{x})$ is small. This test procedure is named Pereira-Stern or full Bayesian significance test (FBST, for short). For more details on Bayesian procedure for test see [18], [12], [21], [22], [17], [4] among many others.

Jeffreys test and the FBST are both Bayes tests for specific loss functions, thus, the decision to be made under both procedures is the action that minimizes the posterior risk ([7], [16]).

Jeffreys and Pereira-Stern measures of evidence are both usefull posterior summaries and, in general, they lead to the same decision. However, some previous works ([20], [17], [15], for example) have shown that decisions that are made taking into consideration such measures can differ. Different decisions are expected whenever the null hypothesis is precise and improper priors or conjugate priors with variance going to infinity are elicited to describe the prior uncertainty about the parameter. Under such priors, Jeffreys test can lead to the Jeffreys-Lindley paradox ([22], [23]) which does not occur if the FBST is assumed [19].

This paper aims at introducing two statistics for testing which consist of pooling Jeffreys and Pereira-Stern measures of evidence in favour of H_0. In order to aggregating these two measures of evidence, we consider the linear and the logarithmic operators which have been widely used in Group Decision Theory in order to obtain a consensus probability measure. Thus, these proposed procedures are intermediate measures of evidence in favour of H_0. We verify the existence of a loss function which render decision theoretic aspects to the proposed procedure which is building assuming the linear operator, that is, we prove that it is also a Bayes test. We also prove that this procedure does not lead to the Jeffreys-Lindley paradox. Although we could not prove that the procedure constructed taking into consideration the logarithmic operator is a Bayes test, we verify that it is a generalization of Jeffreys test. All four procedures are applyed to some asymmetric family of distributions, including the standard skew-normal.

This paper is organized as follows. Section 2 presents two usual mathematical methods for combining or aggregating probability distributions, the linear and the logarithmic operators, and some of their properties. Section 3 breafly presents Jeffreys test and the FBST as a Bayes test. The connection between the FBST and the highest posterior density regions is also provided. Two statistics for testing are introduced and some of their properties are pointed out. In Section 5 the proposed procedures are applied to exponential and standard skew-normal distributions. In order to evaluate the efficiency of the proposed procedures a Monte Carlo study is performed. In Section 6 we test the returns of some Latinamerican emerging markets for asymmetry. Section 7 closes the paper with some conclusions.

2 Pooling probabilities

Combination or aggregation of probabilities plays an important role in decision problems in which there is a group of experts expressing their opinions about the events of interest. This subject has been attracted attention in the literature for many years and many pooling procedures have been proposed in order to obtain the group consensus probability distribution. Two typical and well-known procedures for pooling probabilities are the linear and the logarithm operators. These two aggregating procedures will be, briefly, reviewed in the following. More details about them and some others procedures for pooling probabilities can be found in [8], [9] and [10], and, more recently, a discussion is presented in the context of risk analysis by [6].

Denote by $p_i(\theta)$, $i = 1, \ldots, n$, the opinion of the ith expert about θ which can be a mass function in the discrete case or a density function for the continuous case. Let α_i, $i = 1, \ldots, n$, be non-negative weights such that $\sum_{i=1}^{n} \alpha_i = 1$. The consensus probability distribution P_{Li} is obtained by the *linear probability pool* whenever it is given by:

$$P_{Li}(\theta) = \sum_{i=1}^{n} \alpha_i p_i(\theta). \tag{1}$$

Consider the same notation but, now, assume that $\alpha_i > 0$, $i = 1, \ldots, n$. We say that the consensus probability distribution P_{Lo} is obtained through the *logarithmic probability pool* if it is of form

$$P_{Lo}(\theta) = \frac{\prod_{i=1}^{n} [p_i(\theta)]^{\alpha_i}}{\int_{\Theta} \prod_{i=1}^{n} [p_i(\theta)]^{\alpha_i}} d\theta. \tag{2}$$

These two pooling procedures satisfy some nice properties. The linear probability pool given in (1) preserves unanimity, that is, $P_{Li}(\theta) = a$ if $p_i(\theta) = a$, for all i. Consequently, it satisfies the zero preservation property only if all experts unanimously declare $p_i(\theta) = 0$. On the other hand, it preserves independency only if the group is dictatorial, say, $\alpha_i = 1$ for some i, and expert i announce that the events of interest are independent.

The logarithmic probability pool in (2) also satisfies independency and zero preservation properties. However, it is not necessary unanimity for observing zero preservation property. In fact, such property follows whenever only one expert elicit $p_i(\theta) = 0$. If we assume $\sum_{i=1}^{n} \alpha_i = 1$, the logarithmic probability pool also follows the axiom of unanimity and, under this condition for the weights, the external Bayesianity property is also satisfied, which means that, receiving extra information relevant to θ after $p_i(\theta)$, $i = 1, \ldots, n$, has been declared, the new consensus probability obtained by updating the original one is the same we obtain if we firstly update each expert opinion $p_i(\theta)$ and then combine them.

Next section, we consider these two procedures for aggregating probabilities to obtain statistics for testing.

3 Test procedures

In this section we briefly review Jeffreys and the full Bayesian significance tests. Since Jeffreys and Pereira-Stern measures of evidence in favour of the null hypothesis H_0 are probability measures and can lead to different decisions, we consider the two pooling procedures presented in Section 2 to introduce intermediate measures of evidence in favour of H_0. We prove that the test procedure obtained assuming the linear pool of probability is a Bayes test and does not lead to the Jeffreys-Lindley paradox.

Suppose that we are interested in testing the following hypothesis for θ:

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1, \tag{3}$$

where $\{\Theta_0, \Theta_1\}$ is a partition of Θ, the parametric space of θ.

We focus our attention in tests for precise null hypotheses on θ, that is, we assume $\Theta_0 = \{\theta_0\} \subset \Theta$, where θ_0 is a known value.

3.1 Jeffreys and Pereira-Stern tests

3.1.1 Jeffreys test

Typically, from Bayesian point-of-view, we elicit prior probabilities $P(H_i)$ for the hypotheses $H_i, i = 0, 1$, and compute the posterior probability of H_i through Bayes's theorem. For precise null hypotesis, the posterior for H_0 is given in terms of the Bayes factor $BF(H_0, H_1) = f(\mathbf{x}|H_0)/f(\mathbf{x}|H_1) = BF(H_1, H_0)^{-1}$, say,

$$P(H_0|\mathbf{x}) = \left[1 + \frac{P(H_1)}{P(H_0)} BF(H_1, H_0)\right]^{-1}. \tag{4}$$

We accept H_0 whenever its posterior probability is larger than the posterior probability of H_1. In fact, Jeffreys test is a Bayes test whenever the following loss function is assumed:

$$\begin{cases} \text{L(Accept H}_0, \theta) = \omega_1 \mathbf{1}\{\theta \in \Theta_1\} \\ \text{L(Reject H}_0, \theta) = \omega_0 \mathbf{1}\{\theta \in \Theta_0\}, \end{cases} \tag{5}$$

where $\mathbf{1}\{A\}$ is the indicator function of event A and $\omega_i > 0$, $i = 1, 2$. Thus, we decide in favour of H_0 if the posterior risk of accepting the null hypothesis is the smallest. Consequently, under this appraoch, we accept H_0 whenever

$$P(H_0|\mathbf{x}) > \frac{\omega_1}{\omega_1 + \omega_0}. \tag{6}$$

For a detailed explanation of Jeffreys test see [11], [5], [18] and many others. It is well known that, for testing precise hypothesis, Jeffreys test can lead to the Jeffreys-Lindley paradox ([23], [22]). The FBST, which is briefly described in next section, was introduced in literature in order to avoid such a problem.

3.1.2 Pereira-Stern test

Pereira-Stern or the Full Bayesian significance test (FBST) does not introduce prior probabilities for the hypotheses H_i and makes the test for precise hypotheses simple ([19], [20]). To perform the FBST, the only necessary information is the posterior distribution for θ. In this case, H_0 is accepted if Θ_0 is in a high posterior probability region of Θ.

Let $\pi(\theta|\mathbf{x})$ be the posterior density of θ. Consider the following highest relative surprise (HRS) set:

$$T(\mathbf{x}) = \left\{\theta \in \Theta : \pi(\theta|\mathbf{x}) \geq \sup_{\Theta_0} \{\pi(\theta|\mathbf{x})\}\right\}. \tag{7}$$

The posterior evidence in favour of the null hypothesis is given by $EV(H_0, \mathbf{x}) = 1 - Pr(\theta \in T(\mathbf{x})|\mathbf{x})$. The null hypothesis H_0 is accepted whenever $EV(H_0, \mathbf{x})$ is large.

(See [17] for the FBST in its invariant formulation). As proved in [16], the FBST is a Bayes test if the following loss function is assumed

$$\begin{cases} \text{L}(\text{Accept H}_0, \theta) = b + c\mathbf{1}\{\theta \in T(\mathbf{x})\} \\ \text{L}(\text{Reject H}_0, \theta) = a[1 - \mathbf{1}\{\theta \in T(\mathbf{x})\}], \end{cases} \tag{8}$$

where b, ξ and c are real, positive numbers. Consequently, from the decision theoretic point-of-view, we accept H_0 whenever

$$EV(H_0, \mathbf{x}) > \frac{b+c}{c+a}. \tag{9}$$

Remark: A $100(1-\alpha)\%$ region of highest posterior density (HPD region) for θ is the set $R(\mathbf{x}) = \{\theta \in \Theta : \pi(\theta|\mathbf{x}) \geq c_\alpha\}$ where c_α is the largest constant such that $P(\theta \in R(\mathbf{x})|\mathbf{x}) \geq 1 - \alpha$. It is usual to accept the null hypothesis if the value of θ under test - say, θ_0 - belongs to $R(\mathbf{x})$ [18]. Consequently, decisions made considering the Pereira-Stern measure of evidence and the HPD region are the same whenever

(i) $\alpha < (b+c)(c+a)^{-1} < EV(H_0, \mathbf{x})$ or $(b+c)(c+a)^{-1} < \alpha < EV(H_0, \mathbf{x})$, which leads to the acceptance of H_0;
(ii) $(b+c)(c+a)^{-1} > \alpha > EV(H_0, \mathbf{x})$ or $\alpha > (b+c)(c+a)^{-1} > EV(H_0, \mathbf{x})$, which leads to the rejection of H_0.

Otherwise, Pereira-Stern procedure and the HPD region will lead to different decisions.

3.2 Proposed procedures for testing

Since Jeffreys and Pereira-Stern measures of evidence in favour of the null hypothesis H_0 are probability measures, we can consider them as the opinions of two different experts about the same event and combining them in order to obtain a consensus probability measures that provide an intermediate evidence in favour of H_0.

Considering the linear probability pool given in (1) we have a new measure of evidence in favour of the null hypothesis that is given by:

$$P_{Li}(H_0|\mathbf{x}) = \alpha Ev(H_0, \mathbf{x}) + (1-\alpha)P(H_0|\mathbf{x}), \tag{10}$$

where $\alpha \in [0, 1]$. We decide in favour of the null hypothesis H_0 whenever $P_{Li}(H_0|\mathbf{x})$ is large. This procedure is named along this paper *Linear-pool-based test.*

Considering the logarithmic probability pool in (2), another consensus measure of evidence in favour of H_0 is obtained and assumes the following form:

$$P_{Lo}(H_0|\mathbf{x}) = \frac{[Ev(H_0, \mathbf{x})]^\alpha [P(H_0|\mathbf{x})]^{1-\alpha}}{[Ev(H_0, \mathbf{x})]^\alpha [P(H_0|x)]^{1-\alpha} + [1 - Ev(H_0, \mathbf{x})]^\alpha [1 - P(H_0|\mathbf{x})]^{1-\alpha}}, \tag{11}$$

where $\alpha \in [0, 1]$. Similarly, we decide in favour of the null hypothesis H_0 whenever $P_{Lo}(H_0|\mathbf{x})$ is large. We name this procedure *Logarithmic-pool-based test.* Although it is not necessary, in (11) we assume $\alpha \in [0, 1]$ because, under this condition, the logarithmic probability pool follows the unanimity property.

After some calculations, the probability in (11) become:

$$P_{Lo}(H_0|\mathbf{x}) = \left\{1 + \left[\frac{1 - Ev(H_0, \mathbf{x})}{Ev(H_0, \mathbf{x})}\right]^\alpha \left[\frac{P(H_1)}{P(H_0)} BF(H_1, H_0)\right]^{1-\alpha}\right\}^{-1}, \tag{12}$$

which is a generalization of expression in (4). It is noteworthy that for the non trivial case where $\alpha \neq 0$, if $Ev(H_0, \mathbf{x}) \neq 0$ (even for $Ev(H_0, \mathbf{x})$ very close of 0 which is a strong evidence against the null hypothesis) and $P(H_0|\mathbf{x}) \to 1$ we have that $P_{Lo}(H_0|\mathbf{x}) \to 1$. Similar result is observed for $Ev(H_0, \mathbf{x}) \to 1$ and $P(H_0|\mathbf{x}) \neq 0$.

It is well known that Jeffreys test can lead to the Jeffreys-Lindley paradox ([13], [23]) which states that $P(H_0|\mathbf{x}) \to 1$ for a precise null hypothesis and large sample sizes. This can also be observed if conjugate prior with variance going to infinity or improper priors are assumed [22]. Thus, the procedure in (11) can also lead to the Jeffreys-Lindley paradox since it is enough having $P(H_0|\mathbf{x}) \to 1$ to observe $P_{Lo}(H_0|\mathbf{x}) \to 1$. On the other hand, notice from (10) that $P_{Li}(H_0|\mathbf{x}) \to 1$ only if Jeffreys and Pereira-Stern measure of evidence in favour of H_0 tend both to 1.

Another important characteristic of the Linear-pool-based test in (10) is that it is a Bayes test. In next section, we verify the existence of a loss function that confer a decision theoretic aspect to such a procedure.

3.3 The Bayesianity of the Linear-pool-based test

Let us assume the following loss function:

$$\begin{cases} \text{L(Accept H}_0, \theta) = (1-\alpha)\gamma 1(\theta \in \Theta_1) + \alpha[\beta + \gamma 1(\theta \in T(x))] \\ \text{L(Reject H}_0, \theta) = (1-\alpha)\xi 1(\theta \in \Theta_0) + \alpha\xi[1 - 1(\theta \in T(x))], \end{cases} \tag{13}$$

where $\alpha \in [0,1]$, $\beta \geq 0$, ξ and γ are real, positive numbers.

Theorem: *Minimization of the posterior expected loss function in (13) is the Linear-pool-based test.*

Proof: The posterior risk of accepting H_0 is

$$\begin{aligned} E_\pi(L(\text{Accept } H_0, \theta)|x = & \int_\Theta [(1-\alpha)\gamma 1(\theta \epsilon \Theta_1) + \alpha[\beta + \gamma 1(\theta \epsilon T(x))]]\pi(\theta|x)d\theta \\ & = (1-\alpha)\gamma \int_{\Theta_1} \pi(\theta|x)d\theta + \alpha b \int_\Theta \pi(\theta|x)d\theta + \alpha\gamma \int_{T(x)} \pi(\theta|x)d\theta \\ & = (1-\alpha)\gamma P(H_1|x) + \alpha\beta + \alpha\gamma(1 - Ev(H_0, x)) \\ & = (1-\alpha)\gamma + \alpha(\beta + \gamma) - \gamma P_{LI}(H_0|x). \end{aligned}$$

The posterior risk of rejection is

$$\begin{aligned} E_\pi(L(\text{Reject } H_0, \theta)|x) = & \int_\Theta [(1-\alpha)\xi 1(\theta \epsilon \Theta_0) + \alpha\xi[1 - 1(T(x))]]\pi(\theta|x)d\theta \\ & = (1-\alpha)\xi \int_{\Theta_0} \pi(\theta|x)d\theta + \alpha\xi \int_\Theta \pi(\theta|x)d\theta - \alpha\xi \int_{T(x)} \pi(\theta|x)d\theta \\ & = (1-\alpha)\xi P(H_0|x) + \alpha\xi Ev(H_0, x)) \\ & = \xi P_{LI}(H_0|x). \end{aligned}$$

Therefore, the test is to accept the null hypothesis if, and only if, $E_\pi(L(Accept H_0, \theta)|\mathbf{x}) < E_\pi(L(Reject H_0, \theta)|\mathbf{x})$, that is, whenever we have that:

$$\begin{aligned} (1-\alpha)\gamma + \alpha(b+\gamma) - \gamma P_{LI}(H_0|x) &< \xi P_{LI}(H_0|x) \\ P_{LI}(H_0|x) &> \frac{\gamma + \alpha\beta}{\gamma + \xi}. \end{aligned} \tag{14}$$

4 Criterion to compare Bayes tests

Next sections will present comparisons among the test procedures discussed in the previous one. Since the usual Bayesian procedures for testing and the test proceduce based on the linear aggregation of the usual measures of evidence in favour of the null hypothesis (linear-pool-based test) are Bayes test, in order to fairly compare such procedures, we assume that the prior risks of accepting (rejecting) the null hypothesis are equals for all three procedures and, thus, we define the cut points for acceptance given by (6), (9) and (14).

Denote respectively by $Ev(H_0)$ and $P_{LI}(H_0)$ the prior evidences in favour of the null hypothesis provided by the FBST and the test built using the linear operator. Assume that $Ev(H_0) \in (0,1)$ and $P_{LI}(H_0) \in (0,1)$. By doing equal the prior risks of acceptance for the three procedures, it follows that:

$$c = \frac{\omega_1 P(H_1) - b}{1 - Ev(H_0)}, \tag{15}$$

$$\gamma = \frac{\omega_1 P(H_1) - \alpha\beta}{1 - P_{LI}(H_0)}. \tag{16}$$

Similarly, assuming that the risks of rejecting H_0 for the three procedures are equal, the values of a and ξ are given, respectively, by:

$$a = \frac{\omega_0 P(H_0)}{Ev(H_0)}, \tag{17}$$

$$\gamma = \frac{\omega_0 P(H_0) - \alpha\beta}{P_{LI}(H_0)}. \tag{18}$$

The cut points are then defined by specifying b, β, ω_0 and ω_1. It is noticeable that if $Ev(H_0) = 1$, the constant c is arbitrarely chosen and the expressions to obtain b and a are simplified.

Since we have not found a loss function which render the test constructed using the logarithmic operator (logarithmic-pool-based test), we can assume the same cut point as in (14) whenever it is possible to assume $P(H_0) = Ev(H_0) = p$. Because the unanimity property, in this cases, logarithmic and linear operators provide equal measures of evidence in favour of H_0. Consequently, whenever $p \neq 0$ we have that $\omega_0 = a = \xi$ and the other values are obtained as before. On the other hand, whenever $p = 0$ it follows that $\omega_0 = a = \xi$, $b = \beta = 0$ and c and γ are arbitrarily chosen.

5 Comparing the test procedures

In this section we performe a Monte Carlo study in order to compare the test procedures presented in the previous section. We focus the attention in tests for a precise null hypothesis, say, we assume $H_0 : \theta = \theta_0$, where θ_0 is a known value, and consider two asymmetric families of probability distributions: the exponential and the standard skew-normal distributions. In both cases, we assume $\alpha = 0.5$. We consider two sample sizes ($n = 10$ and 100). We also consider 1,000 replications of the likelihood with parameter θ_{True}.

In order to make the procedures comparable, the cut points k for acceptance of H_0 are defined considering the criterion described in Section 4. We assume that b and β

in expressions (8) and (13) are close to zero. We also assume that $P(H_0)$ is equal or very close to $EV(H_0)$. Under such conditions linear and logarithmic probability pools provide similar results then we can use the same cut point to make decisions about the null hypothesis. Thus, for all test procedures, we will accept the null hypothesis if the posterior evidence in favour of H_0 is higher than $k = \omega_1(\omega_1 + \omega_0)^{-1}$.

5.1 Tests under the exponential distribution

In this section the goal is to evaluate the performance of the proposed procedures whenever a non informative conjugate prior is assumed. We consider a simple situation where $X_1, \ldots, X_n|\theta \overset{iid}{\sim} \exp(\theta)$, $\theta > 0$ and $\theta \sim \text{Gamma}(\psi, \beta)$. Consequently, we have that $\theta|\mathbf{x} \sim \text{Gamma}(\psi + n, \beta + \sum_{i=1}^n x_i)$ and it follows that the Bayes factor and the Pereira-Stern measure of evidence in favour of H_0 are given, respectively, by:

$$FB(H_0, H_1) = \frac{\Gamma(\psi)(\beta + \Sigma_{i=1}^n x_i)^{\psi+n}}{\beta^\psi \Gamma(\psi + n)} \theta_0 e^{-\theta_0 \sum_{i=1}^n x_i},$$

$$Ev(H_0, \mathbf{x}) = 1 - \int_{T(\mathbf{x})} \frac{e^{-\theta(\beta + \sum_{i=1}^n x_i)} \theta^{\psi+n-1} (\beta + \sum_{i=1}^n x_i)^{\psi+n}}{\Gamma(\psi + n)} d\theta, \qquad (19)$$

where $T(\mathbf{x}) = \{\theta \in R_+ : (\psi + n - 1)\log(\theta/\theta_0) \geq (\beta + n\bar{x})(\theta - \theta_0)\}$. Assume that θ_0 is smaller than the posterior mode. Since the Gamma distribution is unimodal, $T(\mathbf{x}) = \{\theta : \theta_0 \leq \theta \leq a\}$, where a is such that $\log(\theta_0) - \theta_0(\beta + \sum_{i=1}^n x_i)[\psi + n - 1]^{-1} = \log(a) - a(\beta + \sum_{i=1}^n x_i)[\psi + n - 1]^{-1}$. Thus, denoting by $\Gamma_a(\alpha, \delta)$ the cdf of the Gamma distribution $\text{Gamma}(\alpha, \delta)$ evaluated in a it follows that:

$$Ev(H_0, \mathbf{x}) = 1 - \Gamma_a \left(\psi + n, \beta + \sum_{i=1}^n x_i \right) + \Gamma_{\theta_0} \left(\psi + n, \beta + \sum_{i=1}^n x_i \right).$$

We obtain $Ev(H_0, \mathbf{x})$ for θ_0 greater than the posterior mode similarly.

In Tables 1 and 2, we provide the average of the posterior measures of evidence in favour of the null hypothesis and the percentage of acceptance of H_0 for different cut points k, respectively. We consider $\theta_0 = 1$ and assume that, *a priori*, $\theta \sim \text{Gamma}(0.001, 0.001)$. Consequently, we have that $EV(H_0) = 0.0063$. Notice from Table 1 that, in average, for both sample sizes, the posterior evidence in favour of H_0 whenever the null hypothesis is true is higher than we have a prior, for all procedures. Moreover, Jeffreys test provides the highest average for the posterior evidence in favour of H_0. The proposed procedures are comparable and provides better result than we obtain for the FBST. For $\theta_{True} \neq \theta_0$ and $n = 10$, in average, the FBST has the best performance always providing the smallest average for the evidence in favour of H_0 and the proposed procedures brings some improvement to the analysis if compared to the Jeffreys test. In this case, the logarithmic-pool-based test is slightly better than the one built assuming the linear probability pool. Similar conclusions can be drawn for $n = 100$. In this case, however, it is noticeable that all four procedures provide the same mean evidence in favour of the null hypothesis, except for $\theta_{True} = 0.9, 1.0$ and 1.5.

It is noteworthy from Table 2 that, for samples of sizes $n = 10$, the percentage of acceptance of H_0 tends to be smaller for high cut points for all procedures. For

Table 1 Posterior mean evidences in favour of H_0, exponential case

θ_{True}	$Ev(H_0, \mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$
Sample size $n = 10$				
0.5	0.0892	0.3982	0.2437	0.1948
0.9	0.4334	0.7990	0.6162	0.6145
1.0	0.5229	0.8151	0.6690	0.6733
1.5	0.4001	0.6837	0.5419	0.5360
2.0	0.1829	0.4715	0.3272	0.3055
2.5	0.0783	0.2960	0.1872	0.1615
3.0	0.0334	0.1614	0.0974	0.0773
Sample size $n = 100$				
0.5	0.0000	0.0000	0.0000	0.0000
0.9	0.3253	0.8495	0.5874	0.5969
1.0	0.5020	0.9228	0.7124	0.7500
1.5	0.0055	0.1156	0.0606	0.0291
2.0	0.0000	0.0000	0.0000	0.0000
2.5	0.0000	0.0000	0.0000	0.0000
3.0	0.0000	0.0000	0.0000	0.0000

$k = 0.90$ and 0.95 the percentage of acceptance of the null hypothesis is very small for all values of θ_{True} (including $\theta_{True} = \theta_0$) and all test procedures and it is zero for the Jeffreys test whenever $k = 0.90$, and for Jeffreys test and the test procedure based on the linear operator whenever $k = 0.95$. For $k = 0.05$ and 0.10, the percentage of acceptance of H_0 tends to be high for all values of θ_{True} (including $\theta_{True} \neq \theta_0$). For the intermediate values of k, the percentage of acceptance tends to be close to zero for values of θ_{True} close to θ_0. The percentage of acceptance under all test procedures is higher then 50.0% for $k = 0.33$ and $\theta_{True} = 0.9$, 1.0 and 1.5. The same result is observed for the Jeffreys test if $k = 0.67$. For $n = 100$ we observe an improvement in the results for all procedures maily for small values of k, that is, assuming the cut points $k = 0.05$ and 0.10 and, for the Jeffreys test whenever $k = 0.90$ and 0.95. It is nothworthy that the FBST presents better performance for θ_{True} close to θ_0. See also Figure 1 for the percentage of rejection of H_0 or the empirical power function of the tests for some particular cases.

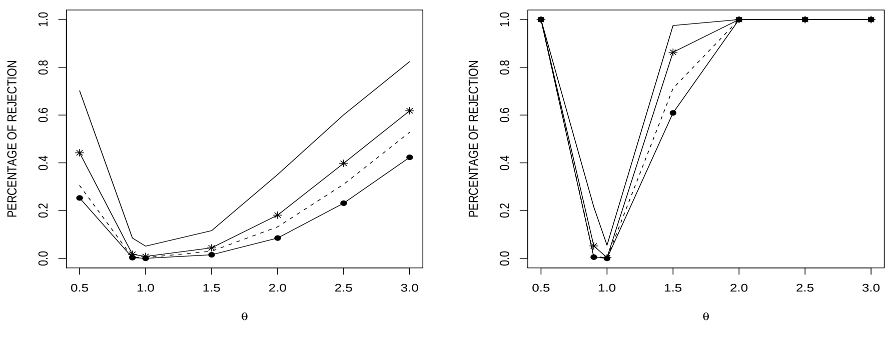

Fig. 1 Emprirical power function for Pereira and Stern (full line), Jeffreys(•), linear-pool-based (dashed line) and logarithimc-pool-based (∗) tests, cut point $k = 0.05$, $n = 10$ (left) and 100, exponential case.

Table 2 Percentage of acceptance of H_0, exponential case

θ_{True}	$Ev(H_0.\mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$	$Ev(H_0.\mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$
	Sample size $n = 10$							
	$k = 0.05$				$k = 0.10$			
0.5	0.298	0.747	0.695	0.558	0.213	0.692	0.620	0.464
0.9	0.915	0.997	0.995	0.983	0.845	0.994	0.989	0.964
1.0	0.949	1.000	0.997	0.992	0.901	0.997	0.991	0.979
1.5	0.884	0.985	0.970	0.956	0.791	0.966	0.945	0.915
2.0	0.649	0.915	0.868	0.819	0.486	0.859	0.787	0.714
2.5	0.399	0.769	0.690	0.602	0.245	0.675	0.571	0.471
3.0	0.176	0.577	0.472	0.382	0.078	0.450	0.318	0.232
	$k = 0.33$				$k = 0.67$			
0.5	0.090	0.531	0.335	0.236	0.028	0.285	0.062	0.081
0.9	0.562	0.975	0.917	0.858	0.251	0.880	0.426	0.482
1.0	0.704	0.979	0.939	0.901	0.374	0.900	0.563	0.629
1.5	0.513	0.891	0.769	0.734	0.237	0.674	0.368	0.404
2.0	0.192	0.656	0.455	0.400	0.049	0.327	0.091	0.105
2.5	0.041	0.404	0.215	0.173	0.005	0.116	0.017	0.021
3.0	0.012	0.178	0.066	0.055	0.002	0.037	0.003	0.003
	$k = 0.90$				$k = 0.95$			
0.5	0.009	0.000	0.007	0.007	0.004	0.000	0.000	0.001
0.9	0.092	0.000	0.076	0.080	0.047	0.000	0.000	0.023
1.0	0.108	0.000	0.089	0.090	0.057	0.000	0.000	0.021
1.5	0.059	0.000	0.045	0.048	0.025	0.000	0.000	0.013
2.0	0.015	0.000	0.012	0.013	0.008	0.000	0.000	0.004
2.5	0.003	0.000	0.003	0.003	0.003	0.000	0.000	0.002
3.0	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000
	Sample size $n = 100$							
	$k = 0.05$				$k = 0.1$			
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.9	0.785	0.995	0.990	0.948	0.687	0.990	0.977	0.912
1.0	0.945	1.000	0.999	0.997	0.882	0.999	0.999	0.993
1.5	0.025	0.391	0.288	0.137	0.011	0.286	0.204	0.069
2.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$k = 0.33$				$k = 0.67$			
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.9	0.417	0.955	0.883	0.784	0.165	0.875	0.366	0.524
1.0	0.662	0.997	0.979	0.938	0.348	0.976	0.597	0.753
1.5	0.001	0.128	0.042	0.018	0.000	0.037	0.001	0.002
2.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$k = 0.90$				$k = 0.95$			
0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.9	0.044	0.645	0.069	0.106	0.018	0.352	0.022	0.024
1.0	0.119	0.833	0.177	0.240	0.058	0.549	0.069	0.072
1.5	0.000	0.004	0.000	0.000	0.000	0.001	0.000	0.000
2.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

In summary, if compared to Jeffreys test, the proposed test procedures are better whenever $\theta_{True} \neq \theta_0$. For $\theta_{True} = \theta_0$, they tend to have better performance than the FBST. The test based on the logarithmic operator is better than the test costructed assuming the linear operator whenever $\theta_{True} \neq \theta_0$ and for k up to 0.33. For such cut points the test based on the linear operator is better if the null hypothesis is true.

5.2 Tests under the skew-normal distribution

Bayesian inference for the skewness parameter λ in the skew-normal family of distribution ([2]) has been considered, for instance, by [14], [3] and [1]. Jeffreys test for λ was firstly considered by [3] that assume two centered student-t prior distributions for λ, with small degrees of freedom - one of which shows to be a good approximation for the reference prior introduced by [14]. In this section, we consider tests for the skewness parameter under the standard skew normal distribution [2] assuming normal priors for the skewness parameter.

Denote by $\phi_n(.;\mu,\Sigma)$ and $\Phi_n(.;\mu,\Sigma)$ ($\phi_n(.)$ and $\Phi_n(.)$) the pdf and the cdf, respectively, of the n-variate normal distribution $N_n(\boldsymbol{\mu},\boldsymbol{\Sigma})$ ($N_n(\mathbf{0},I_n)$). Supose that, given the skewness parameter $\lambda \in R$, the random variables $X_1,...,X_n$ are iid with standard skew-normal distribution [2] which density is $f_{\mathbf{x}}(\mathbf{x}|\lambda) = 2^n\phi_n(\mathbf{x})\Phi_n(\lambda\mathbf{x})$. Assume that $\lambda \sim N(m,v)$. Thus, the posterior is also a skewed distribution which pdf is $\pi(\lambda|\mathbf{x}) = \phi(\lambda;m,v)\Phi_n(\lambda\mathbf{x})\ [\Phi_n(m\mathbf{x};\mathbf{0},I_n+v\mathbf{x}\mathbf{x}^t)]^{-1}$.

Under such assumptions, for testing $H_0 : \lambda = \lambda_0$, we have that the Bayes factor and the Pereira-Stern measure of evidence in favour of H_0 are given, respectively, by:

$$FB(H_0,H_1) = \frac{\Phi_n(\lambda_0\mathbf{x})}{\Phi_n(m\mathbf{x};\mathbf{0},I_n+v\mathbf{x}\mathbf{x}^t)}, \tag{20}$$

$$Ev(H_0,\mathbf{x}) = 1 - \int_{T(\mathbf{x})} \phi(\lambda;m,v)\frac{\Phi_n(\lambda\mathbf{x})}{\Phi_n(m\mathbf{x};\mathbf{0},I_n+v\mathbf{x}\mathbf{x}^t)}d\lambda, \tag{21}$$

where $T(\mathbf{x}) = \{\lambda \in R : \phi(\lambda;m,v)\Phi_n(\lambda\mathbf{x}) \geq \phi(\lambda_0;m,v)\Phi_n(\lambda_0\mathbf{x})\}$.

In the Monte Carlo study in the following, we assume $\lambda_0 = 0$, that is we are testing for normality, and assume prior distributions centered in $m = 0$. As a consequence, the tangencial set is given by $T(\mathbf{x}) = \{\lambda \in R : \lambda^2 \leq 2v\sum_{i=1}^n \log[2\Phi(\lambda x_i)]\}$.

We assume two priors for λ - $\lambda \sim N(0,1)$ and $\lambda \sim N(0,50)$. Thus the prior evidence in favour of H_0 is $EV(H_0) = 1.0$. To define the cut points we also assume $P(H_0) = 0.95$.

Table 3 shows the average of the posterior measures of evidence in favour of the null hypothesis. It can be noticed that, in average, the posterior evidences in favour of H_0 is smaller than we have *a priori* for all procedures, except for Jeffreys test if the null hypothesis is true. Taking, for instance, $n = 10$ and the cut point $k = 0.33$, and assuming that $\lambda \sim N(0,1)$, we conclude that, in average, Jeffreys test leads to the wrong decision for all values of $\lambda_{True} \neq 0$ and the proposed procedures leads to it for values of λ_{True} close to zero.

Table 4 shows the percentage of acceptance of H_0 for different cut points k. See also Figure 2 for the empirical power function for some particular cases. From Table 4 we conclude that all test procedures have better performance for large sample sizes, except whenever $\lambda_{True} = 0$, for the Logarithmic-pool-based test, if $k = 0.67$ and $\lambda \sim N(0,50)$, and for Pereira-Stern and Linear-pool-based tests. Similar to what was observed for the exponential case, Jeffreys test is the best if the null hypothesis is true and Pereira-Stern test works better whenever the null hypothesis is false. For samples of size $n = 100$, the use of an informative prior tends to lead to better results for all test procedures. However, that is not the case for small samples. For $n = 10$, the empirical power function of Jeffreys test tends to be closer to the ideal one for both cut points and if $\lambda \sim N(0,50)$. Similar behavior is observed for the proposed procedures and assuming $k = 0.33$. For $n = 100$ and both prior specifications, its noticeable the good performance of the Logarithmic-pool-based test. The percentage of acceptance of

Table 3 Posterior mean evidences in favour of H_0, skew-normal case

λ_{True}	$Ev(H_0, \mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$
		$\lambda \sim N(0,1)$ and $n = 10$		
-10	0.0107	0.4637	0.2372	0.0915
-5	0.0142	0.5005	0.2574	0.1113
-2	0.0404	0.6198	0.3301	0.2095
-0.5	0.3380	0.9136	0.6258	0.6648
0	0.5065	0.9550	0.7307	0.8010
0.5	0.3281	0.9096	0.6188	0.6552
2	0.0426	0.6316	0.3371	0.2191
5	0.0145	0.5076	0.2611	0.1137
10	0.0107	0.4674	0.2391	0.0923
		$\lambda \sim N(0,50)$ and $n = 10$		
-10	0.0009	0.1181	0.0595	0.0111
-5	0.0022	0.2161	0.1092	0.0299
-2	0.0266	0.5482	0.2874	0.1857
-0.5	0.3063	0.9422	0.6243	0.7255
0	0.4879	0.9826	0.7353	0.8701
0.5	0.3101	0.9342	0.6222	0.7264
2	0.0238	0.5326	0.2782	0.1737
5	0.0025	0.2144	0.1085	0.0310
10	0.0009	0.1147	0.0578	0.0113
		$\lambda \sim N(0,1)$ and $n = 100$		
-10	0.0087	0.0000	0.0043	0.0000
-5	0.0119	0.0000	0.0060	0.0000
-2	0.0118	0.0000	0.0059	0.0000
-0.5	0.0136	0.2927	0.1532	0.0704
0	0.5019	0.9841	0.7430	0.8589
0.5	0.0146	0.2819	0.1483	0.0711
2	0.0119	0.0000	0.0060	0.0000
5	0.0107	0.0000	0.0053	0.0000
10	0.0080	0.0000	0.0040	0.0000
		$\lambda \sim N(0,50)$ and $n = 100$		
-10	0.0003	0.0000	0.0002	0.0000
-5	0.0015	0.0000	0.0008	0.0000
-2	0.0230	0.0000	0.0115	0.0000
-0.5	0.2159	0.5027	0.3593	0.2085
0	0.4875	0.9964	0.7419	0.8381
0.5	0.2269	0.4528	0.3399	0.1964
2	0.0208	0.0000	0.0104	0.0000
5	0.0014	0.0000	0.0007	0.0000
10	0.0004	0.0000	0.0002	0.0000

the null hypothesis for both cut points is close to the one observed for Jeffreys test, when H_0 is true, and it is very close of zero, otherwise.

It is noteworthy that the results obtained for Jeffreys test and the FBST whenever $\lambda \sim N(0,1)$ is comparable to the ones obtained by [3] using the Bayes factor and the 95% HPD region, respectively.

6 Application: Latin American Emerging markets returns

It is well known that emerging markets are more susceptible to the political scenario than developed markets thus their indexes tend to present more atypical observations as well as structural changes. In this cases, normality can be a strong assumption since

Table 4 Percentage of acceptance of H_0, skew-normal case

λ_{True}	$Ev(H_0.\mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$	$Ev(H_0.\mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$
	Sample size $n = 10$							
	$\lambda \sim N(0,1)$							
	$k = 0.33$				$k = 0.67$			
-10	0.000	0.799	0.133	0.010	0.000	0.091	0.000	0.000
-5	0.000	0.847	0.198	0.035	0.000	0.152	0.000	0.000
-2	0.010	0.912	0.478	0.212	0.001	0.439	0.008	0.034
-0.5	0.401	0.995	0.963	0.868	0.173	0.959	0.360	0.587
0	0.686	1.000	0.991	0.961	0.329	0.990	0.651	0.826
0.5	0.405	0.996	0.951	0.857	0.163	0.944	0.372	0.556
2	0.014	0.896	0.512	0.234	0.001	0.477	0.010	0.030
5	0.000	0.845	0.212	0.031	0.000	0.167	0.000	0.001
10	0.000	0.810	0.130	0.007	0.000	0.090	0.000	0.000
	$\lambda \sim N(0,50)$							
	$k = 0.33$				$k = 0.67$			
-10	0.000	0.046	0.002	0.000	0.000	0.002	0.000	0.000
-5	0.000	0.213	0.072	0.003	0.000	0.070	0.000	0.000
-2	0.015	0.648	0.483	0.217	0.002	0.468	0.014	0.057
-0.5	0.373	0.976	0.953	0.870	0.170	0.953	0.363	0.692
0	0.655	0.996	0.990	0.970	0.316	0.990	0.645	0.902
0.5	0.378	0.965	0.944	0.862	0.170	0.940	0.364	0.706
2	0.009	0.652	0.459	0.194	0.005	0.446	0.009	0.059
5	0.000	0.219	0.068	0.009	0.000	0.062	0.000	0.001
10	0.000	0.042	0.004	0.001	0.000	0.004	0.000	0.000
	Sample size $n = 100$							
	$\lambda \sim N(0,1)$							
	$k = 0.33$				$k = 0.67$			
-10	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.5	0.002	0.367	0.194	0.064	0.000	0.188	0.001	0.011
0	0.673	1.000	0.999	0.980	0.340	0.999	0.649	0.899
0.5	0.001	0.344	0.192	0.063	0.000	0.186	0.001	0.013
2	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	$\lambda \sim N(0,50)$							
	$k = 0.33$				$k = 0.67$			
-10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-2	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-0.5	0.312	0.578	0.601	0.265	0.099	0.442	0.039	0.088
0	0.640	1.000	1.000	0.886	0.342	0.999	0.630	0.870
0.5	0.337	0.514	0.558	0.241	0.102	0.384	0.029	0.076
2	0.005	0.000	0.001	0.000	0.001	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

the empirical distributions of such markets often exhibits skewness and tails that are lighter or heavier than normal distribution.

In this section, we consider the return series of the main stock indexes of four Latin American markets, say, the MERVAL (*Índice de Mercado de Valores de Buenos Aires*) of Argentina, the IBOVESPA (*Índice da Bolsa de Valores do Estado de São Paulo*) of Brazil, the IPSA (*Índice de Precios Selectivos de Acciones*) of Chile and the IPyC (*Índice de Precios y Cotizaciones*) of Mexico. The stock returns are recorded weekly from October 31, 1995 to October 31, 2000.

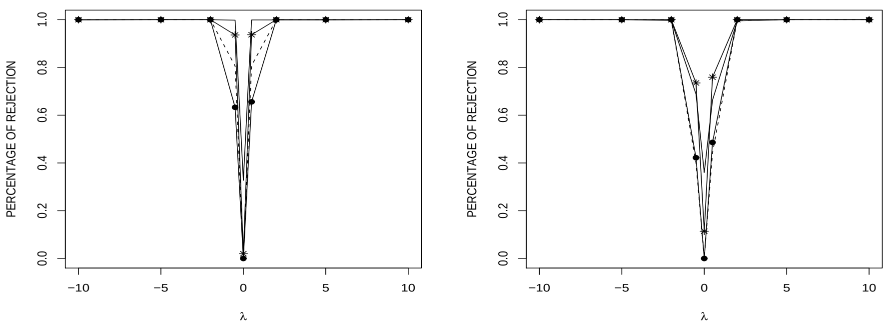

Fig. 2 Emprirical power function for Pereira and Stern (full line), Jeffreys(•), linear-pool-based (dashed line) and logarithimc-pool-based (∗) tests, cut point $k = 0.33$, $n = 100$, $v = 1$(left) and 50, skew-normal case.

The data sets consist of the return series $r_1, \ldots, r_n$ that were transformed in order to be better fitted by the standard skew-normal distribution with unknown skewness parameter, say, the data sets are $y_i = r_i(\bar{r^2})^{-0.5}$, $i = 1, \ldots, n$, where $\bar{r^2} = \sum_{i=1}^{n} r_i^2/n$. (By using this tranformation it follows that $E(Y) = 1$ and CV(Y)=CV(R)).

We assume that $\lambda \sim N(0, 50)$, consequently, the prior evidence in favour of H_0 is 1.0. We also consider two prior specifications for H_0 - a non-informative prior which establishes that $P(H_0) = 0.5$ and the other one that assume that $P(H_0)$ is close to the prior Pereira-Stern measure of evidence in favour of the null hypothesis, that is, we assume $P(H_0) = 0.99$. Under the last prior, we can assume the same cut point k for accepting H_0 for all procedures, since in this case the linear and the logarithmic probability pool have similar behavior.

Table 5 and Figure 3 present, respectively, posterior summaries and the densities for the skewness parameters for the four indexes. The area in grey represents the posterior Pereira-Stern measure of evidence in favour of the null hypothesis.

Table 5 Posterior summaries for the skewness parameter

	mean	variance	mode
MERVAL	0.0016	0.0004	0.0838
IBOVESPA	0.0982	0.0130	0.0986
IPSA	-0.0008	0.0003	-0.0455
IPyC	0.1338	0.0137	0.1436

Notice from Figure 3 that the posteriors of λ for all indexes have unique modes and put most of their mass in small values of λ which means that we have evidence of small asymmetry for all stock returns. It can also be perceived from Table 5. Notice that Bayes estimates (posterior means and modes) point out that the skewness parameters λ for all indexes are small. Moreover, for MERVAL and IPSA, the estimates for λ are very close of zero which means that, for Argentinean and Chilean stock markets, the assumption of normality for the returns can be a reazonable one. IPyC presents the

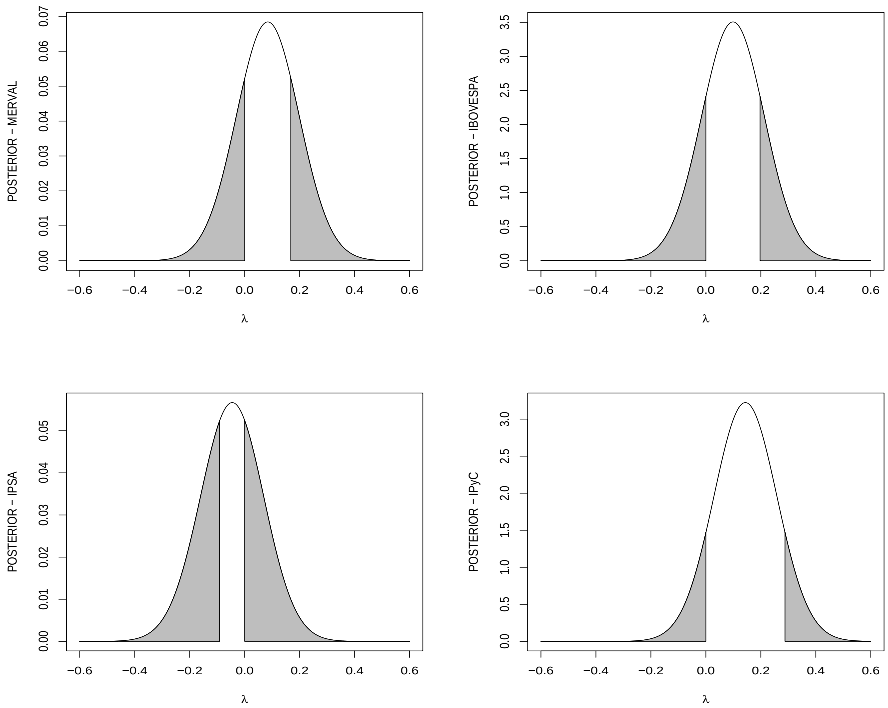

Fig. 3 Posteriors for the skewness parameters and posterior evidences $EV(H_0, x)$ (area in grey)

highest asymmetry - for this index, we observe $\hat{\lambda}$ close to 0.13. The asymmetry for all indexes is positive, except for IPSA.

Table 6 Tests for the skewness parameter

		$P(H_0) = 0.99$			$P(H_0) = 0.50$		
	$EV(Ho, \mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$	$P(H_0\|\mathbf{x})$	$P_{Li}(H_0\|\mathbf{x})$	$P_{Lo}(H_0\|\mathbf{x})$
MERVAL	0.9895	0.9892	0.9894	0.9894	0.4808	0.7352	0.9033
IBOVESPA	0.3869	0.9998	0.6933	0.9810	0.9771	0.6820	0.8384
IPSA	0.9950	0.9892	0.9921	0.9927	0.4815	0.7383	0.9315
IPyC	0.2676	0.9996	0.6336	0.9685	0.9630	0.6153	0.7551

Table 6 presents the posterior evidences in favour of H_0 for all procedures. Notice that the proposed procedures and Pereira-Stern posterior evidences in favour of H_0 are higher for MERVAL and IPSA than for IBOVESPA and IPyC, for all choices of $P(H_0)$ while the opposite is observed for Jeffreys test. We also notice that logarithmic and Jeffreys test provide very close evidence in favour of H_0, mainly, for $P(H_0) = 0.99$.

For $P(H_0) = 0.99$ and assuming that $\omega_0 = \omega_1$ (say, $k = 0.50$), Pereira-Stern test leads to the conclusion that the returns of MERVAL and IPSA are symmetric and that the returns of IBOVESPA and IPyC are asymmetric. All the other procedures lead to the conclusion that the returns for all indexes are distributed according to the standard normal distribution. If we assume that $2\omega_0 = \omega_1$ the same conclusion is drawn, except that in this case the Linear-pool-based test also points out that the distribution of the IPyC returns has asymmetric behaviour. Jeffreys and Logarithmic-pool-based tests provide strong evidence in favour of the null hypothesis in all cases - it is noteworthy that we will accept the null hypothesis for all cut points $k < 0.9685$. Notice also that Pereira-Stern and the Linear-pool-based tests lead to the same conclusion for all $k > 0.6933$ and $k < 0.9894$, in this cases, accepting the null hypothesis only for MERVAL and IPyC.

Now asssume that $P(H_0) = 0.50$. In this case the evidences in favour of the null hypothesis provided by the Linear-pool-based and Logarithmic-pool-based tests are higher than 0.61 for all indexes, which means that we have a relatively strong evidence of asymmetry for the returns. Consider an arbitrary cut point $k = 0.90$ which is an usual cut point in classical significance test procedures. In this case, the FBST and the Logarithmic-pool-based test lead to the conclusion that the returns of IPSA and MERVAL have symmetric behavior, that is, they are normally distributed. The same conclusion is drawn for the returns of IPSA, IBOVESPA, IPyC and MERVAL if the Linear-pool-based test is assumed as the test procedure. Moreover, in this case, Pereira-Stern and Jeffreys tests lead to opposite decisions. Similar behaviour is observed when comparing Jeffreys and Logarithmic-pool-based tests. Notice that according to Jeffreys test we conclude that MERVAL and IPSA has asymmetric behaviour what is a contradiction if we take into consideration the estimates in Table 5.

7 Final Remarks

In this paper we introduced two Bayesian procedures for hypotheses testing which are based on aggregating Jeffreys and Pereira-Stern measures of evidence in favour of the null hypothesis. These procedures were constructed considering the linear and the logarithmic operators which are typical procedures to obtain a consensus probability in Group Decision Theory. We performed a Monte Carlo study in order to compare all the four procedures assuming asymmetric families of distribution. We applied the procedures to test the returns of some Latin American emerging stock markets for asymmetry.

From the simulation study we concluded that, in general, the proposed test procedures tend to be better than Jeffreys test whenever the null hypothesis is false, and they tend to have better performance than Pereira-Stern test (FBST) whenever the null hypothesis is true. The Logarithmic-pool-based (Linear-pool-based) test tends to be better than the linear-pool-based (logarithmic-pool-based) one whenever the null hypothesis is false and small (large) cut points are considered.

Overall, the proposed procedures, mainly the Logarithmic-pool-based test, bring some improvement and show to be reazonable approaches for testing.

Acknowledgements The authors would like to express their gratitude to Ricardo C. Takahashi and Gustavo M. A. Rocha (UFMG-Brazil) for comments and suggestions in earlier versions of this paper. R. H. Loschi and C. C. Santos acknowledges CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil, grants 304505/2006-4 (RHL) and 502419/2007-5 (CCS), for a partial allowance to her research. The research of R. B. Arellano-Valle was supported in part by FONDECYT (Chile), grant 1085241.

References

1. Arellano-Valle, R. B, Genton, M. G., Loschi, R. H., Shape mixtures of multivariate skew-normal distributions. Journal of Multivariate Analysis **100**, 91–101 (2008).
2. Azzalini, A., A class of distributions which includes the normal ones, Scand. J. Statist. **12**, 171–178 (1985).
3. Bayes, C. L. and Branco, M. B., Bayesian Inference for the Skewness Parameter of the Scalar Skew-Normal Distribution, Brazilian Journal of Probability and Statistics, **21** (2), 141–163 (2007).
4. Berger, J. O. and Delampady, M., Testing precise hypotheses, Statistical Science **2**, 317–352 (1987).
5. Bernardo, J. M. and Smith, A.F.M., Bayesian Theory(1st edn). Wiley, Chichester, (1994).
6. Clemen, R. T. and Winkler, R. L., Combining probability distributions from experts in risk analysis, Risk Analysis, **19** (2), 187–203 (1999).
7. DeGroot M. H., Probability and Statistics(2nd edn). Addison Wesley, USA, (1989).
8. French, S., Group consensus probability distributions: A critical survey, Bayesian Statistics,Bernardo, J. M. ,DeGroot M. H., Lindley, D. V. and Smith, A.F.M. (eds), **2**, 183–202 (1985).
9. Genest, C. and Zidek, J. V., Combining probability distributions: A critique and an annoted bibliography, Statistical Science, **1** (1), 114–148 (1986).
10. Genest, C. McConway, K.J. and Schervish, M. J., Characterization of externally Bayesian pooling operators, The Annals of Statistics, **14** (2), 487–501 (1986).
11. Jeffreys, H., Theory of Probability. Oxford, Claredon Press, (1961).
12. Lavine, M. and Schervish, M. J., Bayes Factors: What they are and what they are not, The American Statistician, **53**(2), 119–122 (1999).
13. Lindley, D. V., A statistical paradox, Biometrika, **44**(1/2), 197–192 (1957).
14. Liseo,b., Loperfido, N., A note on reference priors for the scalar skew-noraml distribution, Journal of Statistical Planning and Inference, **136**, 373–389 (2006).
15. Loschi R.H., Monteiro, J. V. D., Rocha,G. H. M. A. and Mayrink, V. D. . Testing and estimating the non-disjunction fraction in meiosis I using reference priors, Biometrical Journal **49**(6),824–839 (2007).
16. Madruga M.R., Esteves L.G., Wechsler S., On the Bayesianity of Pereira-Stern tests, Test, **10**(1), 291–299 (2001).
17. Madruga M.R., Pereira C.A.B., Stern, J. M., Bayesian evidency test for precise hypotheses. Journal of Statistical Planning and Inference, **117**, 185–198 (2003).
18. Migon H.S. and Gamerman D., Statistical Inference: An integrated approach. Arnold, New York, USA (1999).
19. Pereira C.A.B. and Stern J.M., Evidence and credibility: Full Bayesian significance test for precise hypotheses. Entropy **1**, 69-80 (1999).
20. Pereira C.A.B. and Stern J.M., Model selection: Full Bayesian approach, Envirometrics **12**, 559-568 (2001).
21. Pereira C.A.B., Stern J.M. and Wechsler, S., Can a significance test be genuinely Bayesian?, Bayesian Analysis, **3** (1), 79-100 (2008).
22. Robert, C. P., A note on Jeffreys-Lindley paradox, Statistica Sinica, **3**, 601–608 (1993).
23. Tsao, C. A., A note on Lindley's paradox, Test **15**(1), 125–139 (2006).