Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

COMPARING HIERARCHICAL MODELS USING BAYES FACTOR AND FRACTIONAL BAYES FACTOR: A ROBUST ANALYSIS

BY FULVIO DE SANTIS AND FULVIO SPEZZAFERRI
Carnegie Mellon University and Universita' di Roma "La Sapienza"

In the recent years several alternative Bayes Factors have been introduced in order to handle the problem of the extreme sensitivity of the Bayes Factor (BF) to the priors of the models under comparison in model selection or hypothesis testing problems. In particular, the impossibility of using the Bayes Factor with standard noninformative priors has led to introduce new automatic criteria as the Intrinsic Bayes Factors (IBFs) and the Fractional Bayes Factor (FBF). As pointed out by De Santis-Spezzaferri (1995), the use of IBFs and of the FBF seems to be appealing also in robust Bayesian analyses, when the priors of the parameters of the models vary in large classes of distributions, containing, in the limiting case, improper priors. In this paper we study the behaviour of the BF and of the FBF in a problem of comparing two hierarchical models. We assume the exchangeability of the parameters and introduce a class of distributions at the third stage of the hierarchy of the "biggest" model. In this context, the use of the FBF seems to avoid the problems of lack of robustness of the BF, providing an alternative to the use of the BF itself.

1. Introduction. Suppose that we want to compare two models M_1 and M_2 given some data y. Let $f_i(y|\eta_i)$ and $\pi_i(\eta_i)$ be respectively the likelihood and the prior distribution of model M_i, $i = 1, 2$. A measure of the evidence given by the data y in favour of model M_2 versus M_1 is represented by the *Bayes factor* (BF) that is defined as

$$B_{21}(y) = \frac{m_2(y)}{m_1(y)}$$

where $m_i(y) = \int f_i(y|\eta_i)\pi_i(\eta_i)d\eta_i$, is the marginal distribution of the data y under model M_i, $i = 1, 2$.

The BF is extremely sensitive to prior assumptions since it depends on the absolute values of the priors of the parameters. Specifically, several problems arise when prior information is weak (see for example Aitkin, 1991, O'Hagan, 1995 and De Santis-Spezzaferri, 1995, for a discussion on this topic). In fact the use of reference priors is not possible, since they are typically improper and hence defined only up to arbitrary constants that do not cancel out in the resulting BF. A possible solution to this problem is to split the sample into two parts $y(l)$ and $y(n-l)$ and then to use $y(l)$ as a *training sample* to convert improper priors into proper ones, and the rest of the data to compute the BF, called *Partial Bayes factor* (PBF):

$$B_{21}(l) = \frac{\int f_2(y(n-l)|\eta_2)\pi(\eta_2|y(l))d\eta_2}{\int f_1(y(n-l)|\eta_1)\pi(\eta_1|y(l))d\eta_1}.$$

where $\pi(\eta_i|y(l))$ is the posterior distribution of the parameter η_i given $y(l)$, $i = 1, 2$. The *PBF*, whose expression can easily be written as $B_{21}(l) = B_{21}(y)/B_{21}(y(l))$ does not depend on arbitrary constants anymore.

In order to eliminate the dependence of the *PBF* on the particular training sample $y(l)$ several methods have been lately proposed. Among the others, the most interesting seem to be the *Intrinsic Bayes factors* (Berger and Pericchi, 1994, 1995 and 1996) and the *Fractional Bayes factor* (O'Hagan, 1995).

Berger and Pericchi suggest several appropriate averages of the *PBF* over the set of all possible training samples and to choose l so that $y(l)$ is a *minimal training sample*, that is a subsample of minimal size such that $0 < m_i(y(l)) < \infty$, $i = 1, 2$.

O'Hagan's method is based on the consideration that, when both l and n are large, the likelihood based on the subsample $y(l)$ approximates the likelihood based on the whole sample raised to the power $h = l/n$ and so he proposes to replace $B_{21}(y(l))$ in the denominator of the *PBF* with the expression $B^l_{21}(y) = \int f^h_2(y|\eta_2)\pi_2(\eta_2)d\eta_2/\int f^h_1(y|\eta_1)\pi_1(\eta_1)d\eta_1$. In the nested model case, De Santis and Spezzaferri (1995) show that, under suitable conditions on the sample distributions, the *FBF* is actually a *PBF* if the training sample leads to the same maximum likelihood estimators found using the whole sample y. Therefore the *FBF* is a particular *PBF* that uses as training sample $y(l)$ a, regular say, subsample containing, if it exists, the same information of the whole sample y. Further interpretations of *FBF* are given in Wasserman (1994), Kass-Wasserman (1995) and Pauler (1995).

Closely related to the sensitivity of the *BF* to prior assumptions is the lack of robustness of the *BF* when the priors vary in classes that contain diffuse distributions (see for example Berger-Delampady, 1987, Berger-Sellke, 1987 and Sanso'-Pericchi-Moreno, 1996). In this case, for nested models, $\inf B_{21}(y) \approx 0$, regardless of the observed data y, making it impossible a robust choice of M_2. On the contrary, the correction term in *FBF* has the property of balancing the penalization of M_2 in correspondence of flat priors, leading therefore to lower bounds that are dependent on the data and, in general, not uniformly equal to zero (see De Santis-Spezzaferri, 1995). Sanso'-Moreno-Pericchi (1996) and De Santis-Spezzaferri (1995) point out that a similar feature is also shared by a particular version of *IBF*, that is the *expected arithmetic IBF* (see Berger-Pericchi, 1996, for a definition). This property suggests the use of *FBF* and of the *expected arithmetic IBF* in robust analyses.

In this paper we compare two nested normal hierarchical models, using *BF* and *FBF* and introducing suitable classes of priors for the parameters of the highest level of the hierarchy (see also Berger-Pericchi, 1994, for comparisons of hierarchical models). We show that, while *BF* presents the usual non-robust behaviour, FBF leads to lower and upper bounds that can be used for model selection purposes.

2. Model and prior assumptions. Let us now specify the models introduced in the previous Section. Suppose that $\eta_2 = (\theta_2, \sigma^2)$ and $\eta_1 = (\theta_1, \sigma^2)$ and that M_i, $i = 1, 2$, are two nested normal linear models, such that

$$y|\theta_i, \sigma^2 \sim N(X_i\theta_i, \sigma^2 I_n)$$

where y is a vector on n observations, X_i is the design matrix of order (n, k_i), θ_i is a k_i-parameter vector and σ^2 is in general unknown.

We assume that $X_2 = (X_1 \vdots X)$ and $\theta_2^T = (\theta_1^T \vdots \theta^T)$ with θ a k-vector, X a (n, k) matrix and $k_2 = k_1 + k$. Without loss of generality, we also assume that the columns of X are orthogonal to those of X_1.

Let us now turn to the prior specifications. We assume the conditional independence of θ_1 and θ under M_2 and the same distribution for the common parameters (θ_1, σ^2)

$$\pi_2(\theta_2, \sigma^2) = \pi(\theta_1|\sigma^2)\pi(\theta|\sigma^2)\pi(\sigma^2) \quad \text{and} \quad \pi_1(\theta_1, \sigma^2) = \pi(\theta_1|\sigma^2)\pi(\sigma^2).$$

For the parameter vector θ_1 we assign an exchangeable prior distribution. Specifically, we assume that, given the hyperparameter μ

$$\theta_1|\mu, \sigma^2 \sim N(\mu 1_{k_1}, c\sigma^2 I_{k_1}) \text{ with } c > 0$$

and, for μ, the reference prior $\pi(\mu) \propto cost.$ Observe that under the above assumptions on the common parameter θ_1, using the reference prior for μ does not leave the resulting BF undetermined. In fact the reference prior for μ depends on an arbitrary constant that cancels out in the resulting BF, since it appears in both the numerator and the denominator of BF itself.

Let us now focus on the parameter vector θ, and assume that its distribution varies in a class of exchangeable priors. In the following we will consider two different classes for $\pi(\theta|\sigma^2)$ which, for semplicity's sake will be simply denoted by $\pi(\theta)$. The first class Γ_1 is

$$\Gamma_1 = \{\pi(\theta) : \ \pi(\theta) = \int \pi(\theta|\alpha)p(\alpha)d\alpha, \ p(\alpha) \in \Gamma_{US}\}$$

where

$$\theta|\alpha \sim N(\alpha 1_k, c\sigma^2 I_k)$$

and Γ_{US} = {all unimodal and symmetric distributions with mode zero}. Assuming a class of distributions on the location hyperarameter α, we introduce different levels of dependence between the components of the vector θ. As a limiting case we obtain the independence when $p(\alpha)$ assigns probability one to the point zero. In general it can be shown that the correlation coefficient between two components of θ is $var(\alpha)/(c\sigma^2 + var(\alpha))$. Therefore, more diffuse $p(\alpha)$ is, higher the positive correlations among the elements of θ are. The class Γ_{US} is typically used for a location parameter (see for example Berger-Delampady, 1987, and Berger-Sellke, 1987) and it presents the sensible feature that the corresponding class Γ_1 contains distributions that spread out the mass around the central value zero without bias towards specific alternatives.

A different way to formalize the prior information on θ is obtained considering the class

$$\Gamma_2 = \{\pi(\theta) : \pi(\theta) = \int \pi(\theta|\gamma)p(\gamma)d\gamma, \quad p(\gamma) \in \Gamma_U\}$$

where

$$\theta|\gamma \sim N(0_k, \gamma\sigma^2 I_k) \qquad \gamma > 0$$

and $\Gamma_U = \{$all unimodal distributions with mode in zero$\}$. Differently from the distributions previously considered, the components of the vector θ are uncorrelated when $\pi(\theta) \in \Gamma_2$. Analogously to Γ_1, the independence is achieved as a limiting case when $p(\gamma)$ assigns probability one to the point zero. The components of θ are in general stochastically dependent and the conditional variance of θ_h given θ_k is an increasing function of the absolute value of θ_k, where $(\theta_h,\ \theta_k)$ are two generic elements of θ. This can be easily proved from the equality $var(\theta_h|\theta_k) = \sigma^2 E(\gamma|\theta_k)$ observing that the conditional random variable $\gamma|\theta_k$ is stochastically increasing in $|\theta_k|$. The class Γ_U is a natural candidate for a scale parameter (see for example Berger-Mortera, 1994). It is possible, anyway, to consider other classes of distributions for γ, such as all the Gamma densities with mode different from zero. We observe however that Γ_U contains, in the limiting case, the distribution that gives identical predictive distributions under the two models. Therefore Γ_U is a sensible choice when we want to compare M_1 to arbitrarily close alternative models.

Let us focus more on the different dependence structures for the components of θ when $\pi(\theta)$ belongs to Γ_1 or Γ_2. Regardless of specific choices for the classes $p(\alpha)$ and $p(\gamma)$ vary in, the conditional mean of θ_h given θ_k is an increasing function of θ_k if $\pi(\theta) \in \Gamma_1$ while the conditional mean of $|\theta_h|$ given θ_k is an increasing function of $|\theta_k|$ if $\pi(\theta) \in \Gamma_2$.

Finally note that using both Γ_1 and Γ_2, we reduce the multidimensional robustness problem to a unidimensional one.

3. Bounds for the Bayes factor and the Fractional Bayes factor. In order to evaluate the robustness of BF and FBF with respect to the prior assumptions introduced in Section 2, we first consider, for a fixed

c, the bounds of BF and FBF, when $\pi(\theta) \in \Gamma_1$ or Γ_2. Then, using the famous Hald data set, we numerically evaluate the bounds for several values of c.

3.1. *General results for known σ^2.* Let us first consider the simple case of a known σ^2. Under the model and prior assumptions previously described, BF and FBF, regardless of $\pi(\theta_1|\sigma^2)$, are respectively

$$B_{21}(y) = exp\{\frac{y^T y}{2\sigma^2}\} \int exp\{-\frac{1}{2\sigma^2}(y - X\theta)^T(y - X\theta)\}\pi(\theta)d\theta$$

and $B_{21}^F(l) = B_{21}(y)/B_{21}^l(y)$ where

$$B_{21}^l(y) = exp\{h\frac{y^T y}{2\sigma^2}\} \int exp\{-\frac{h}{2\sigma^2}(y - X\theta)^T(y - X\theta)\}\pi(\theta)d\theta.$$

RESULT 1. *Given c and using the class Γ_1 the upper bounds of BF and FBF are respectively*

$$zep \sup_{\pi(\theta)\in\Gamma_1} B_{21}(y) = K \sup_{z>0} \frac{1}{2z}\{\Phi(\sqrt{a}(z - \frac{b}{a\sigma})) - \Phi(\sqrt{a}(-z - \frac{b}{a\sigma}))\}$$

$$\sup_{\pi(\theta)\in\Gamma_1} B_{21}^F(l) = \frac{K}{K(h)} \sup_{z>0} \frac{\Phi(\sqrt{a}(z - \frac{b}{a\sigma})) - \Phi(\sqrt{a}(-z - \frac{b}{a\sigma}))}{\Phi(\sqrt{a(h)h}(z - \frac{b(h)}{a(h)\sigma})) - \Phi(\sqrt{a(h)h}(-z - \frac{b(h)}{a(h)\sigma}))}$$

where $\Phi(.)$ is the cumulative distribution function (c.d.f.) of the standard normal distribution,

$$K = \sqrt{\frac{2\pi}{a}} \frac{exp\{\frac{1}{2\sigma^2}(y^T H y + \frac{b^2}{a})\}}{|cXX^T + I_n|^{1/2}}, \quad a = 1_k^T X^T(I_n - H)X1_k, \quad b = 1_k^T X^T(I_n - H)y$$

$$H = X(X^T X + c^{-1}I_k)^{-1}X^T \quad and \quad K(h) = \sqrt{\frac{2\pi}{a(h)h}} \frac{exp\{\frac{h}{2\sigma^2}(y^T H(h)y + \frac{b^2(h)}{a(h)})\}}{|chXX^T + I_n|^{1/2}}$$

where $a(h)$, $b(h)$ and $H(h)$ are obtained from a, b and H replacing c with ch. Lower bounds are obtained similarly.

Representing $\pi(\theta)$ as a mixture of $\pi(\theta|\alpha)$ for a given $p(\alpha)$, the first part of Result 1 can be derived performing the integration in the above expression of BF first with respect to θ, then with respect to α and then maximizing the resulting functional in $p(\alpha) \in \Gamma_{US}$. The second part of Result 1 is obtained applying the above procedure to both $B_{21}(y)$ and $B_{21}^l(y)$. Observing that BF and FBF are respectively a linear functional and a ratio of linear functionals in $p(\alpha)$, the upper bounds are obtained from Lemma

A.1 in Sivaganesan-Berger (1989) expressing $p(\alpha)$ as a mixture of uniform distributions in $[-z, z]$, $z > 0$.

RESULT 2. *Using the class Γ_2, the upper bounds of BF and FBF are respectively*

$$\sup_{\pi(\theta)\in\Gamma_2} B_{21}(y) = \sup_{z>0} \frac{\psi_2(z)}{z} \quad and \quad \sup_{\pi(\theta)\in\Gamma_2} B_{21}^F(l) = \sup_{z>0} \frac{\psi_2(z)}{\psi_2^h(z)}$$

where

$$\psi_2(z) = \int_0^z \frac{1}{|\gamma X^T X + I_k|^{1/2}} exp\{\frac{1}{2\sigma^2} y^T X (X^T X + \gamma^{-1} I_k)^{-1} X^T y\} d\gamma$$

and $\psi_2^h(z)$ is obtained from $\psi_2(z)$ replacing γ with γh and σ^2 with σ^2/h. Lower bounds are obtained similarly.

Result 2 can be derived similarly to Result 1 expressing $p(\gamma)$ as a mixture of uniform distributions in $[0, z]$, $z > 0$. As a remark, observe that using Γ_2, BF and FBF are independent of c.

EXAMPLE. As an example, we consider the Hald's cement data, described in detail in Draper-Smith (1981) and also analyzed by Berger-Pericchi (1994), using several versions of $IBFs$ and other Bayesian choice criteria. There are $n = 13$ observations and four regressors x_1, x_2, x_3 and x_4. For computational semplicity we orthogonalize and standardize the matrix X_2, whose columns are the values assumed by the four regressors, and we center the response variable in the origin. In this example, M_2 is the full model in which the four regressors are included. We compare M_2 alternatively to model M_1^{12} or to M_1^{34}, where M_1^{ij} denotes the model in which only the regressors x_i and x_j are included. We use the specific version of FBF obtained assuming $l = k_2 = 4$, that is, in this case, the minimal training sample size (see Berger-Mortera, 1995). Bounds for BF and FBF are computed using Γ_1 for several values of σ^2 and c and reported in Table 1 (for $M_1 = M_1^{12}$) and Table 2 (for $M_1 = M_1^{34}$). In order to consider a wide range of evidence in favour of M_1^{12}, we have assigned the values $\sigma^2 = 1, 2, 5$, with corresponding p-values equal to 0.006, 0.08 and 0.37 (the m.l.e. for σ^2 under M_2 and M_1^{12} are respectively 3.68 and 4.45). For the comparison of M_2 and M_1^{34} we consider only the value $\sigma^2 = 100$ and p-value=1.6 10^{-6} (the m.l.e. for σ^2 under M_1^{34} is 207.31), since smaller values give an even stronger evidence against the latter model. As expected, the lower bounds of the BF are uniformly equal to zero (therefore they are omitted) and the upper bounds are highly sensitive to c. The evidence provided by the bounds of the FBF in favour of a model is consistent with the p-values corresponding to the given σ^2. The robust behaviour of FBF is pointed out by the values of its extrema that, for increasing values of c become closer and closer and essentially stable with respect to c itself. For istance (see Table 1), when $\sigma^2 = 1$ (p-value=0.006), FBF leads to a robust choice of M_2, while BF does not.

Similar remarks hold for the bounds of BF and FBF using Γ_2, reported in Table 3 ($M_1 = M_1^{12}$) and Table 4 ($M_1 = M_1^{34}$). As previously mentioned, the bounds of BF and FBF are now independent of c. FBF shows now extrema whose ranges are wider than the ones found using Γ_1. This is because the class Γ_2 contains the degenerate distribution that makes M_2 equal to M_1 and that forces the value one to belong to the interval given by the extrema of FBF.

TABLE 1

Bounds for BF and FBF using Γ_1 ($M_1 = M_1^{12}$)

	$\sigma^2 = 1$			$\sigma^2 = 2$			$\sigma^2 = 5$		
c	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$
0	2.09	1.00	2.19	1.37	1.00	1.24	1.00	0.71	1.00
1	6.99	5.59	6.84	1.75	1.65	1.94	0.83	0.75	1.00
10	8.72	10.37	11.09	0.89	1.82	2.03	0.23	0.67	0.73
100	1.43	10.05	10.16	0.12	1.77	1.79	0.03	0.63	0.62

TABLE 2

Bounds for BF and FBF using Γ_1 ($M_1 = M_1^{34}$)

σ^2	p-value	c	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$
100	1.610^{-6}	1	66793	192	4920

TABLE 3

Bounds for BF and FBF using Γ_2 ($M_1 = M_1^{12}$)

$\sigma^2 = 1$			$\sigma^2 = 2$			$\sigma^2 = 5$		
$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$
9.37	1.00	10.25	1.66	1.00	2.03	1.00	0.63	1.03

TABLE 4

Bounds for BF and FBF using Γ_2 ($M_1 = M_1^{34}$)

σ^2	p-value	$\bar{B}_{21}$	$\underline{B}^F_{21}$	$\bar{B}^F_{21}$
100	1.610^{-6}	12952	1	3047

3.2. *General results for unknown σ^2.* Considering σ^2 as a nuisance parameter, we integrate it out using, under the two models, the standard noninformative prior $\pi(\sigma^2) = 1/\sigma^2$. Different prior assumptions can be considered to express ignorance on σ^2. We shall not attempt to justify the use of the standard noninformative prior in this framework. Some discussion on *well-calibrated* priors for model selection problems can be found in Berger-Pericchi

(1995). In this case BF and FBF are respectively

$$B_{21}(y) = \frac{\int f_2(y|\theta_1, \theta, \sigma^2)\pi(\theta_1|\mu)p(\mu)\pi(\theta)\pi(\sigma^2)d\theta_1 d\mu d\theta d\sigma^2}{\int f_1(y|\theta_1, \sigma^2)\pi(\theta_1|\mu)p(\mu)\pi(\sigma^2)d\theta_1 d\mu d\sigma^2}$$

and $B_{21}^F(l) = B_{21}(y)/B_{21}^l(y)$ where $B_{21}^l(y)$ is obtained from the expression $B_{21}(y)$ replacing σ^2 with σ^2/h in the likelihoods.

RESULT 3. *Given c, using the class Γ_1, the upper bounds of BF and FBF are respectively*

$$\sup_{\pi(\theta)\in\Gamma_1} B_{21}(y) = K' \sup_{z>0} \frac{1}{2z}\{T_{n-1}(z - \frac{b}{as}) - T_{n-1}(-z - \frac{b}{as})\}$$

$$\sup_{\pi(\theta)\in\Gamma_1} B_{21}^F(l) = \frac{K'}{K'(h)} \sup_{z>0} \frac{T_{n-1}(z - \frac{b}{as}) - T_{n-1}(-z - \frac{b}{as})}{T_{l-1}(z - \frac{b(h)}{a(h)s(h)}) - T_{l-1}(-z - \frac{b(h)}{a(h)s(h)})}$$

where $T_i(.)$ is the c.d.f. of the t-distribution with i degrees of freedom,

$$K' = (\frac{y^T W_1^{-1} y}{y^T W_2^{-1} y - \frac{b^2}{a}})^{n/2} \frac{\sqrt{(n-1)\pi}\,\Gamma(\frac{n-1}{2})}{|I_k + cX^T X|^{1/2}\Gamma(\frac{n}{2})}, \qquad s = \sqrt{\frac{y^T W_2^{-1} y - b^2/a}{(n-1)a}},$$

$$W_1^{-1} = I_n - X_1(X_1^T X_1 + c^{-1}(I_{k_1} - (k_1)^{-1} J_{k_1}))^{-1} X_1^T, \quad W_2^{-1} = W_1^{-1} - H,$$

a, b, H, $a(h)$ and $b(h)$ are defined in Result 1, and where J_{k_1} is the (k_1, k_1) unity matrix. Finally, $K'(h)$ and $s(h)$ are obtained from K' and s replacing c with ch and n with l. Lower bounds are obtained similarly.

To obtain Result 3 we first evaluate the standard integrations in the denominator of $B_{21}(y)$. Then, representing $\pi(\theta)$ as a mixture of $\pi(\theta|\alpha)$ for a given $p(\alpha)$, we perform the integrations in the numerator with respect to θ_1, μ, θ, σ^2 and then with respect to α. The same procedure is applied to both $B_{21}(y)$ and $B_{21}^l(y)$ to obtain an explicit expression for $B_{21}^F(l)$. The resulting functionals in $p(\alpha) \in \Gamma_{US}$ are then maximized as in Result 1.

RESULT 4. *Using the class Γ_2, the upper bounds of BF and FBF are respectively*

$$\sup_{\pi(\theta)\in\Gamma_2} B_{21}(y) = \sup_{z>0} \frac{\varphi_2(z)}{z} \quad \textit{and} \quad \sup_{\pi(\theta)\in\Gamma_2} B_{21}^F(l) = \sup_{z>0} \frac{\varphi_2(z)}{\varphi_2^h(z)}$$

where

$$\varphi_2(z) = \int_0^z \frac{1}{|\gamma X^T X + I_k|^{1/2}}(1 - \frac{y^T X(X^T X + \gamma^{-1} I_k)^{-1} X^T y}{y^T W_1^{-1} y})^{-n/2} d\gamma$$

and $\varphi_2^h(z)$ *is obtained from* $\varphi_2(z)$ *replacing* γ *with* γh, c *with* ch *and* n *with* l. *Lower bounds are obtained similarly.*

Result 4 is obtained similarly to Result 3 representing $\pi(\theta)$ as a mixture of $\pi(\theta|\gamma)$ for a given $p(\gamma) \in \Gamma_{US}$.

EXAMPLE (CONTINUED). Table 5 reports the bounds for BF and FBF using the class Γ_1 when σ^2 is unknown and $l = k_2 + 1 = 5$. The lower bounds of BF are not reported since uniformly equal to zero. The p-value of M_1^{12} is now approximately 0.47 and the p-value of M_1^{34} is approximately 0.005. The same remarks for the known σ^2 case hold also in this situation. Again, the bounds of BF are highly sensitive to c, while the bounds of FBF are consistent to the corresponding p-values and essentially stable with respect to c. Finally we observe that the results shown in Table 5 are consistent with those corresponding to the similar p-value in the known σ^2 case and reported in the last column of Table 1 ($\sigma^2 = 5$).

TABLE 5

Bounds for BF and FBF using Γ_1 ($M_1 = M_1^{12}$, $M_1 = M_1^{34}$)

	M_1^{12}			M_1^{34}		
c	$\bar{B}_{21}$	$\underline{B}_{21}^F$	$\bar{B}_{21}^F$	$\bar{B}_{21}$	$\underline{B}_{21}^F$	$\bar{B}_{21}^F$
0	1.00	1.00	1.23	$2.6\ 10^9$	1.00	$4.0\ 10^6$
1	0.88	1.08	1.18	$3.6\ 10^9$	$2.0\ 10^6$	$4.7\ 10^6$
10	0.27	0.89	0.94	$2.4\ 10^9$	$4.4\ 10^5$	$4.2\ 10^6$
100	0.03	0.80	0.85	$6.9\ 10^8$	$2.0\ 10^6$	$3.8\ 10^6$

4. Conclusions. This paper dealt with the robustness of BF and FBF with respect to prior assumptions for a comparison of nested hierchical models. For the common parameter θ_1 we have considered an exchangeable prior depending on the unknown constant c. To evaluate the sensitivity of BF and FBF to the more critical assumptions on θ, we have introduced the classes Γ_1 and Γ_2 that represent different dependence structures for the components of θ. We have also observed that using Γ_1 and Γ_2 reduce the multidimensional robustness problem to a univariate one, and that using Γ_2 allows comparisons with arbitrarily close alternative models. In the Example considered, unlike BF, FBF obtained for l equal to the minimal training sample size has shown a robust behaviour with respect to both Γ_1 and Γ_2. Furthermore the bounds of FBF were essentially stable with respect to c. Therefore FBF seems to be able to lead to robust model selection.

REFERENCES

AITKIN, M. (1991). Posterior Bayes factors (with discussion). *J. Roy. Statist. Soc.* Ser. B, 53 111-142

BERGER, J. O. AND DELAMPADY, M. (1987). Testing precise hypotheses. *Statist. Sci.,* 3 317-352.

BERGER, J. O. AND MORTERA, J. (1994). Robust Bayesian hypothesis testing in the presence of nuisance parameters. *J. Statist. Planning and Inference*, 40, 357-373.

BERGER, J. O. AND MORTERA, J. (1995). In Discussion of O'Hagan. *J. Roy. Statist. Soc.* Ser. B, 57 99-138.

BERGER, J. O. AND SELLKE, T. (1987). Testing a point null hypothesis: the irreconciliability of P-values and evidence. *J. Amer. Statist. Assoc.,* 82 112-122.

BERGER, J. O. AND PERICCHI, L. R. (1994). The Intrinsic Bayes factor for linear models. In *Bayesian Statistics V 23-42*, J. M. Bernardo et al. eds. Oxford University Press.

BERGER, J. O. AND PERICCHI, L. R. (1995). On the justification of default and Intrinsic Bayes factors. Tech. Report 95-18c. Dept of Statistics, Purdue University.

BERGER, J. O. AND PERICCHI, L. R. (1996). The Intrinsic Bayes factor for model selection and prediction. *J. of Amer. Statist. Ass.*, 91, 109–122.

DE SANTIS, F. AND SPEZZAFERRI, F. (1995). Alternative Bayes factors for model selection. Submitted.

DRAPER, N. R. AND SMITH, H. (1981). Applied Regression Analysis. New York: J. Wiley and sons. II edition.

KASS, R.E., AND WASSERMAN, L. (1995). A reference bayesian test for nested hypotheses and its relatioship with the Schwarz criterion. *J. of Amer. Statist. Ass.,* 90, 928–934.

O'HAGAN, A. (1995). Fractional Bayes factors for model comparison (with discussion). *J. Roy. Statist. Soc.* Ser. B, 57 99-138.

PAULER, D. (1995). The Schwarz criterion for statistical models. Thesis proposal, Dep. of Statistics, Carnegie Mellon University.

SANSO', B., PERICCHI, L. R. AND MORENO, E. (1996). On the robustness of the Intrinsic Bayes factor for nested models. *Bayesian Robustness.* IMS Lecture Notes Monograph Series, Vol. 29.

SIVAGANESAN, S. AND BERGER, J. O. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.,* 17 868-889.

WASSERMAN, L. (1994). The conflict between improper priors and robustness. Tech. Report, Depart. of Statistics, Carnegie Mellon University.

CARNEGIE MELLON UNIVERSITY
DEPARTMENT OF STATISTICS
PITTSBURGH, PA 15213

UNIVERSITA' DI ROMA
"LA SAPIENZA"
ROMA, ITALY