Article

Adaptive Significance Levels in Tests for Linear Regression Models: The *e*-Value and *P*-Value Cases

Alejandra E. Patiño Hoyos [1,2,*] ⓘ**, Victor Fossaluza** [2,*] ⓘ**, Luís Gustavo Esteves** [2]
and Carlos Alberto de Bragança Pereira [2] ⓘ

[1]   Facultad de Ingeniería, Institución Universitaria Pascual Bravo, Medellín 050034, Colombia
[2]   Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo 05508-090, Brazil
*   Correspondence: alejandra.patino@pascualbravo.edu.co (A.E.P.H.); victorf@ime.usp.br (V.F.)

Abstract: The full Bayesian significance test (FBST) for precise hypotheses is a Bayesian alternative to the traditional significance tests based on *p*-values. The FBST is characterized by the *e*-value as an evidence index in favor of the null hypothesis (**H**). An important practical issue for the implementation of the FBST is to establish how small the evidence against **H** must be in order to decide for its rejection. In this work, we present a method to find a cutoff value for the *e*-value in the FBST by minimizing the linear combination of the averaged type-I and type-II error probabilities for a given sample size and also for a given dimensionality of the parameter space. Furthermore, we compare our methodology with the results obtained from the test with adaptive significance level, which presents the capital-P *P*-value as a decision-making evidence measure. For this purpose, the scenario of linear regression models with unknown variance under the Bayesian approach is considered.

Keywords: adaptive significance levels; Bayesian test; linear regression; predictive distribution; significance test

1. Introduction

The full Bayesian significance test (FBST) for precise hypotheses is presented in [1] as a Bayesian alternative to the traditional significance tests based on *p*-values. With the FBST, the authors introduce the *e*-value as an evidence index in favor of the null hypothesis (**H**). An important practical issue for the implementation of the FBST is to establish how small the evidence must be to decide to reject **H** ([2,3]). In that sense, the authors of [4] present loss functions such that the minimization of their posterior expected values characterizes the FBST as a Bayes test under a decision-theoretic approach. This procedure provides a cutoff point for the evidence that depends on the severity of the error for deciding whether to reject or accept **H**.

In the frequentist significance-test context, it is known that under certain conditions the *p*-value decreases as the sample size increases, in such a way that by setting a single significance level, the comparison of the *p*-value with the fixed significance level usually leads to rejection of the null hypothesis ([5–9]). In the FBST procedure, the *e*-value exhibits similar behavior to the *p*-value when the sample size increases, which suggests that the cutoff point to define the rejection of **H** should depend on the sample size and (possibly) on other characteristics of the statistical model under consideration. However, in the proposal of [4], a loss function that explicitly takes into account the sample size is not studied.

In order to solve the problem of testing hypotheses in the usual way, in which changing the sample size influences the probability of rejecting or accepting the null hypothesis, the authors of [10], motivated by [11], suggest that the level of significance in hypothesis testing should be a function of sample size. Instead of setting a single level of significance, the authors of [10] propose fixing the ratio of severity between type-I and type-II error probabilities based on the incurred losses in each case, and thus, given a sample size, defining the level of significance that minimizes the linear combination of the decision error

probabilities. The authors of [10] show that proceeding this way, by increasing the sample size, the probabilities of both kind of errors and their linear combination decrease, while in most cases, setting a single level of significance independent of sample size, only type-II error probability decreases. The tests proposed in [10] take the same conceptual grounds of the usual tests for simple hypotheses based on the minimization of a linear combination of probabilities of error of decisions as presented in [12]. Then, the authors of [10] extend, in a sense, the idea in [12] to composite and sharp hypotheses, according to the initial work in [11].

Following the same line of work, the authors of [13,14] present a new hypothesis-testing procedure formulated from the ideas developed in previous works ([11,15–17]) and using a mixture of frequentist and Bayesian tools. This procedure introduces the capital-P *P*-value as a decision-making evidence measure and also includes an adaptive significance level, i.e., a significance level that is a function of sample size. Such an adaptive significance level is obtained from the minimization of the linear combination of generalized type-I and type-II error probabilities. According to the authors of [14], the resulting hypothesis tests do not violate the likelihood principle and do not require any constraints on the dimensionalities of the sample space and parameter space. It should be noticed that the new test procedure is precisely the optimal decision rule for the problem of testing the simple hypotheses $f_{\mathbf{H}}$ against $f_{\mathbf{A}}$. For this reason, such a procedure overcomes the drawback of increasing the sample size resulting in the rejection of a null precise hypothesis ([12]). Another important way of successfully dealing with this question is to take into account meaningful deviations from the parameter value that specifies the null precise hypothesis in the formulation of the hypothesis testing problem ([18,19]).

On the other hand, linear models are probably the most used statistical models to establish the influence of a set of covariates on a response variable. In that sense, the proper identification of the relevant variables in the model is an important issue in any scientific investigation and is a more challenging task in the context of Big-Data problems. In addition to high dimensionality, in recent statistical learning problems, it is common to find large datasets with thousands of observations. This fact may cause the hypothesis of nullity of the regression coefficients to be rejected most of the time, due to the large sample size when the significance level is fixed.

The main goal of our work is to determine, in the setting of linear regression models, how small the Bayesian evidence in the FBST should be in order to reject the null hypothesis and prevent a decision-maker from the abovementioned drawbacks. Therefore, taking into account the concepts in [11,12] associated with optimal hypothesis tests, as well as the conclusions in [10] about the relationship between the significance levels and the sample size, and finally, considering the ideas developed recently by the authors of [13,14] related to adaptive significance levels, we present a method to find a cutoff point for the *e*-value by minimizing a linear combination of the averaged type-I and type-II error probabilities for a given sample size and also for a given dimensionality of the parameter space. For that purpose, the scenario of linear regression models with unknown variance under the Bayesian approach is considered. So, by providing an adaptive level for decision making and controlling the probabilities of both kinds of errors, we intend to avoid the problems associated with the rejection of the hypotheses on the regression coefficients when the sample size is very large. In addition to the *e*-value, we calculate the *P*-value as well as its corresponding adaptive significance levels in order to compare the decisions that can be made by performing the tests with each of these measures.

2. The Linear Regression Model with Unknown Variance

The identification of the relevant variables in linear models can be done through hypothesis-testing procedures involving the respective regression coefficients. In the conjugate Bayesian analysis of the normal linear regression model with unknown variance, it is possible to obtain expressions for the posterior distributions of the parameters and their respective marginals. Therefore, in this setting, the FBST can be used for testing if one or

more of the regression coefficients is null, which is the basis of one possible model-selection procedure. We first review the normal linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \tag{1}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is an $n \times 1$ vector of y_i observations, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ is an $n \times p$ matrix of covariates, also called the design matrix, with $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip-1})^\top$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_p)^\top$ is a $p \times 1$ vector of parameters (regression coefficients), and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ an $n \times 1$ vector of random errors. The model shows simply that the conditional distribution of $\mathbf{y}$ given parameters $(\boldsymbol{\theta}, \sigma^2)$ is the multivariate normal distribution $N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n)$. Therefore, the likelihood becomes

$$f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})\right\}. \tag{2}$$

The natural conjugate prior distribution of $(\boldsymbol{\theta}, \sigma^2)$ is a p-variate normal-inverse gamma distribution with hyperparameters $\mathbf{m_0}$, $\mathbf{V_0}$, a_0, and b_0, denoted by $(\boldsymbol{\theta}, \sigma^2) \sim N_p IG(\mathbf{m_0}, \mathbf{V_0}, a_0, b_0)$. Combining it with the likelihood (2) gives the posterior distribution ([20–22]):

$$f(\boldsymbol{\theta}, \sigma^2|\mathbf{y}) \propto (\sigma^2)^{-\left(a_0 + \frac{n}{2} + \frac{p}{2} + 1\right)} \exp\left\{-\frac{1}{2\sigma^2}\left[(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{V}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*) + 2b_1\right]\right\}, \tag{3}$$

where

$$\mathbf{V}^* = \left(\mathbf{V_0}^{-1} + \mathbf{X}^\top \mathbf{X}\right)^{-1}, \quad \mathbf{m}^* = \mathbf{V}^* \left(\mathbf{V_0}^{-1} \mathbf{m_0} + \mathbf{X}^\top \mathbf{y}\right),$$

$$a_1 = a_0 + \frac{n}{2}, \quad b_1 = b_0 + \frac{\mathbf{m_0}^\top \mathbf{V_0}^{-1} \mathbf{m_0} + \mathbf{y}^\top \mathbf{y} - \mathbf{m}^{*\top} \mathbf{V}^{*-1} \mathbf{m}^*}{2}.$$

If $\mathbf{X}^\top \mathbf{X}$ is non-singular, we can write

$$\mathbf{m}^* = \mathbf{V}^* \left(\mathbf{V_0}^{-1} \mathbf{m_0} + \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\theta}}\right),$$

where $\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is the classical maximum likelihood or least squares estimator of $\boldsymbol{\theta}$. Therefore, the posterior distribution of $(\boldsymbol{\theta}, \sigma^2)$ is

$$(\boldsymbol{\theta}, \sigma^2)|\mathbf{y} \sim N_p IG(\mathbf{m}^*, \mathbf{V}^*, a_1, b_1).$$

See Appendix A for further explanation of the priors, posteriors, and conditional distributions for the linear regression models with unknown variance.

3. Adaptive Significance Levels in Linear Regression Coefficient Hypothesis Testing

In this section, we present the methodology to find a cutoff value for the evidence in the FBST as an adaptive significance level and we also develop the procedure to calculate the *P*-value with its corresponding adaptive significance level, all this in the context of linear regression coefficient hypothesis testing in models with unknown variance under the Bayesian point of view. For that purpose, first of all, it is necessary to show how the Bayesian prior predictive densities under the null and alternative hypotheses are defined.

3.1. Prior Predictive Densities in Regression-Coefficient Hypothesis Testing

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top \boldsymbol{\theta}_2^\top)^\top$, with $\boldsymbol{\theta}_1 = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_s)^\top$ and $\boldsymbol{\theta}_2 = (\boldsymbol{\theta}_{s+1}, \ldots, \boldsymbol{\theta}_p)^\top$, having $\boldsymbol{\theta}_1$ s elements and $\boldsymbol{\theta}_2$ r elements. Let $\boldsymbol{\xi} = (\boldsymbol{\theta}^\top, \sigma^2)^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \sigma^2)^\top$, then, $\mathbf{Y}|\boldsymbol{\xi} \sim N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n)$ where $\boldsymbol{\xi} \in \Xi$. We are interested in testing the hypotheses

$$\mathbf{H} : \boldsymbol{\theta}_2 = \mathbf{0}$$
$$\mathbf{A} : \boldsymbol{\theta}_2 \neq \mathbf{0}.$$

Let $\Xi_{\mathbf{H}}$ and $\Xi_{\mathbf{A}}$ be the partition of the parameter space defined by the competing hypotheses $\mathbf{H}$ and $\mathbf{A}$. Consider the prior density $g(\boldsymbol{\zeta})$ defined over the entire parameter space Ξ and let $f_{\mathbf{H}}$ and $f_{\mathbf{A}}$ be the Bayesian prior predictive densities under the respective hypotheses. Both are probability density functions over the sample space $\boldsymbol{\Omega}$, as follows:

$$f_{\mathbf{H}}(\mathbf{y}) = t_n \left(2\left(a_0 + \frac{r}{2}\right); \mathbf{XCm}_{\mathbf{01.2}}(\mathbf{0}), \frac{b_0 + \dfrac{\mathbf{m_{02}}^\top (\mathbf{V_{022}})^{-1} \mathbf{m_{02}}}{2}}{\left(a_0 + \dfrac{r}{2}\right)} \left(\mathbf{I}_n + (\mathbf{XC}) \mathbf{V_{011.2}} (\mathbf{XC})^\top \right) \right), \quad (4)$$

where $\mathbf{C}_{(s+r) \times s} = [\mathbf{I}_s, \mathbf{0}_{s \times r}]^\top$.

Additionally,

$$f_{\mathbf{A}}(\mathbf{y}) = t_n \left(2a_0; \mathbf{Xm_0}, \frac{b_0}{a_0} \left(\mathbf{I}_n + \mathbf{XV_0X}^\top \right) \right). \quad (5)$$

where $P_{\mathbf{H}}$ and $P_{\mathbf{A}}$ are the prior probability measure of $\boldsymbol{\zeta}$ restricted to the sets $\mathbf{H}$ and $\mathbf{A}$ respectively (more details can be seen in Appendix B).

3.2. Evidence Index: e-Value

The *full Bayesian significance test* (FBST) was proposed in [1] for precise or "sharp" hypotheses (subsets of the parameter space with smaller dimension than the dimension of the whole parameter space, and, therefore, with null Lebesgue measure) based on the evidence in favor of the null hypothesis, calculated as the posterior probability of the complement of the highest posterior density (HPD) region (here we consider the usual HPD region with respect to the Lebesgue measure, even though it could be built by choosing any other dominating measure instead) tangent to the set that defines the null hypothesis. Considering the concepts in [10,11], and the recent works [13,14] related to adaptive significance levels, we propose to establish a cutoff value k^* for the e-value ($ev(\mathbf{H}; \mathbf{y}_0)$) in the FBST as a function of the sample size n and the dimensionality of the parameter space d, i.e., $k^* = k^*(n, d)$ with $k^* \in [0, 1]$, such that k^* minimizes the linear combination of the averaged type-I and type-II error probabilities, $a\alpha + b\beta$. To describe the procedure in the context of the coefficient hypothesis testing of the linear regression model we are addressing, consider the tangential set to the null hypothesis which is defined as

$$\begin{aligned} T_{\mathbf{y}_0} &= \left\{ \boldsymbol{\zeta} \in \Xi : f(\boldsymbol{\zeta}|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\zeta}|\mathbf{y}_0) \right\} \\ &= \left\{ (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \in \Xi : f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\mathbf{y}_0) \right\}. \end{aligned} \quad (6)$$

This is the posterior distribution of $(\boldsymbol{\theta}_1, \sigma^2)$ given $\boldsymbol{\theta}_2$ a s-variate normal-inverse gamma, that is

$$(\boldsymbol{\theta}_1, \sigma^2|\boldsymbol{\theta}_2, \mathbf{y}_0) \sim N_s IG \left(\mathbf{m}^*_{1.2}(\boldsymbol{\theta}_2), \mathbf{V}^*_{11.2}, a_0 + \frac{r}{2}, b_0 + \frac{(\boldsymbol{\theta}_2 - \mathbf{m}^*_2)^\top \mathbf{V}^{*-1}_{22} (\boldsymbol{\theta}_2 - \mathbf{m}^*_2)}{2} \right),$$

where the point under **H** for which the posterior attains its maximum value can be calculated as follows

$$
\begin{aligned}
\arg\sup_{\mathbf{H}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2 | \mathbf{y}_0) &= \arg\sup_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2 | \mathbf{y}_0) \\
&= \arg\sup_{\boldsymbol{\theta}_1, \sigma^2} \frac{f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2 | \mathbf{y}_0)}{\int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s, \sigma^2 \in \mathbb{R}_+} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2 | \mathbf{y}_0) \, d\boldsymbol{\theta}_1 \, d\sigma^2} \\
&= \arg\sup_{\boldsymbol{\theta}_1, \sigma^2} f(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}, \mathbf{y}_0) \\
&= \mathrm{Mode}\left[f(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}, \mathbf{y}_0) \right] \\
&= \left[\mathbf{m}^*_{1.2}(\boldsymbol{\theta}_2 = \mathbf{0}), \mathbf{0}, \frac{b_1 + \dfrac{(\mathbf{m}^*_2)^\top (\mathbf{V}^*_{22})^{-1}(\mathbf{m}^*_2)}{2}}{\left(a_1 + \dfrac{r}{2}\right) + 1 + \dfrac{s}{2}} \right] \\
&= \left[\widehat{\boldsymbol{\theta}}_1, \mathbf{0}, \widehat{\sigma^2} \right].
\end{aligned}
$$

Thus, we get the tangential set

$$
T_{\mathbf{y}_0} = \left\{ (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \in \Xi : f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2 | \mathbf{y}_0) > f(\widehat{\boldsymbol{\theta}}_1, \mathbf{0}, \widehat{\sigma^2} | \mathbf{y}_0) \right\}. \tag{7}
$$

The evidence in favor **H** is calculated as the posterior probability of the complement of $T_{\mathbf{y}_0}$. That is,

$$
ev(\mathbf{H}; \mathbf{y}_0) = 1 - P(\boldsymbol{\xi} \in T_{\mathbf{y}_0} | \mathbf{y}_0). \tag{8}
$$

The evidence index, *e*-value, in favor of a precise hypothesis, considers all points of the parameter space which are less "probable" than some point in $\Xi_{\mathbf{H}}$. A large value of $ev(\mathbf{H}; \mathbf{y}_0)$ means that the subset $\Xi_{\mathbf{H}}$ lies in a high-probability region of Ξ, and, therefore, the data support the null hypothesis; on the other hand, a small value of $ev(\mathbf{H}; \mathbf{y}_0)$ means that $\Xi_{\mathbf{H}}$ is in a low-probability region of Ξ and the data would make us discredit the null hypothesis ([23]).

The evidence in (8) can be approximately determined via Monte Carlo simulation. Then, generating M samples from the posterior distribution of $\boldsymbol{\xi}$, such that $\boldsymbol{\xi} | \mathbf{y} \sim N_p IG(\mathbf{m}^*, \mathbf{V}^*, a_1, b_1)$, we estimate the evidence by Monte Carlo simulation through the expression

$$
1 - \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}\left(\boldsymbol{\xi}^{(j)} \in T_{\mathbf{y}_0} \right).
$$

Now, consider the test such that

$$
\varphi_e(\mathbf{y}) = \begin{cases} 0 & if \quad ev(\mathbf{H}; \mathbf{y}) > k \\[2mm] 1 & if \quad ev(\mathbf{H}; \mathbf{y}) \leq k. \end{cases}
$$

The averaged error probabilities, expressed in terms of the predictive densities, can be estimated by Monte Carlo simulation through the expressions

$$
\alpha_{\varphi_e} = \int_{\mathbf{y} \in \Psi_e} f_{\mathbf{H}}(\mathbf{y}) \, d\mathbf{y} \quad \text{and} \quad \beta_{\varphi_e} = \int_{\mathbf{y} \notin \Psi_e} f_{\mathbf{A}}(\mathbf{y}) \, d\mathbf{y}, \tag{9}
$$

where Ψ_e is the set

$$
\Psi_e = \{ \mathbf{y} \in \boldsymbol{\Omega} : ev(\mathbf{H}; \mathbf{y}) \leq k \}.
$$

So, the adaptive cutoff value k^* for $ev(\mathbf{H}; \mathbf{y})$ will be the k that minimizes $a\alpha_{\varphi_e} + b\beta_{\varphi_e}$. The a and b values represent the relative seriousness of errors of the two types or, equivalently, relative prior preferences for the competing hypotheses. For example, if $b/a = 1$, it is said that β_{φ_e} and α_{φ_e} are equally severe, whereas if $b/a < 1$, then α_{φ_e} undergoes a more intense minimization than β_{φ_e}, which means that type-I error is considered more serious than type-II error and also indicates a prior preference for **H**.

3.3. Significance Index: P-Value

The authors of [13,14] present a new hypothesis-testing procedure using a mixture of frequentist and Bayesian tools. On the one hand, the procedure resembles a frequentist test as it is based on the comparison of the *P*-value as a decision-making evidence measure with an adaptive significance level. On the other hand, such an adaptive significance level is obtained from the minimization of a linear combination of generalized type-I and type-II error probabilities under a Bayesian perspective. As a result, it generally depends on both the null and alternative hypotheses and on the sample size n, as opposed to standard fixed significance levels. The new proposal may also be seen as a test for simple hypotheses characterized by the predictive distributions $f_{\mathbf{H}}$ and $f_{\mathbf{A}}$ in Section 3.1 that minimizes a specific linear combination of probabilities of errors of decision. It is then formally characterized by a cutoff for the Bayes Factor (which takes the place of the likelihood ratio here) and therefore may prevent a decision-maker from rejecting the null hypothesis when the data seem to be clear evidence in its favor ([12]). It should be stressed that under the new proposal, a cutoff value for the Bayes factor (for the "likelihood ratio" here) is chosen in advance and consequently no constraint is imposed exclusively on the probability of the error of the first kind. In this sense, the test in [13,14] completely departs from regular frequentist tests. From another angle, the Bayes factor may be seen as the ratio between the posterior odds in favor of the null hypothesis and its prior odds ([24]). Note that the quantity defined here is a capital-P "*P*-value" to distinguish it from the small-p "*p*-value". In the scenario of the linear regression model with unknown variance, the ratio between the two prior predictive densities (4) and (5), will be the Bayes factor,

$$\mathrm{BF}(x) = \frac{f_{\mathbf{H}}(x)}{f_{\mathbf{A}}(x)}. \tag{10}$$

Now, consider the test

$$\varphi^*(\mathbf{y}) = \begin{cases} 0 & if \quad \mathrm{BF}(\mathbf{y}) > \dfrac{b}{a} \\[2mm] 1 & if \quad \mathrm{BF}(\mathbf{y}) \leq \dfrac{b}{a}. \end{cases}$$

For any other test φ, φ^* minimizes a linear combination of the type-I and type-II error probabilities, $a\alpha_\varphi + b\beta_\varphi$. Here again, the a and b values represent the relative seriousness of errors of the two types. To obtain the *P*-value at the point $\mathbf{y}_0 \in \mathbf{\Omega}$, define the set Ψ_0 of sample points $\mathbf{y}$ for which the Bayes factors are smaller than or equal to the Bayes factor of the observed sample point $\mathbf{y}_0$, that is

$$\Psi_0 = \{\mathbf{y} \in \mathbf{\Omega} : \mathrm{BF}(\mathbf{y}) \leq \mathrm{BF}(\mathbf{y}_0)\}.$$

Then, the *P*-value is the integral of the predictive density over **H**, $f_{\mathbf{H}}$, in Ψ_0

$$P\text{-value}(\mathbf{y}_0) = \int_{\Psi_0} f_{\mathbf{H}}(\mathbf{y}) \, d\mathbf{y}.$$

Defining the set Ψ^* of sample points **y** with Bayes factors smaller than or equal to b/a, i.e.,

$$\Psi^* = \left\{ \mathbf{y} \in \mathbf{\Omega} : \mathrm{BF}(\mathbf{y}) \leq \frac{b}{a} \right\},$$

the optimal averaged error probabilities from the generalized Neyman–Pearson Lemma, which will depend on the sample size, are given by

$$\alpha_{\varphi^*} = \int_{\mathbf{y} \in \Psi^*} f_{\mathbf{H}}(\mathbf{y}) \, d\mathbf{y} \quad \text{and} \quad \beta_{\varphi^*} = \int_{\mathbf{y} \notin \Psi^*} f_{\mathbf{A}}(\mathbf{y}) \, d\mathbf{y}.$$

In order to make a decision, the *P*-value is compared to the optimal adaptive significance level α_{φ^*}. Then, when $\mathbf{y}_0$ is observed, the hypothesis **H** will be rejected if the *P*-value $(\mathbf{y}_0) < \alpha_{\varphi^*}$.

4. Simulation Study

We developed a simulation study considering two models. The first model was

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \tag{11}$$

where $\mathbf{X} = \mathbf{1}_n$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. The hypotheses to be tested were

$$\mathbf{H} : \boldsymbol{\theta}_1 = 0$$
$$\mathbf{A} : \boldsymbol{\theta}_1 \neq 0.$$

The second model studied was

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \tag{12}$$

where $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ is an $n \times p$ matrix of covariates with $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{ip-1})^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is the $p \times 1$ vector of coefficients. In this case, the hypotheses of interest were

$$\mathbf{H} : \boldsymbol{\theta}_2 = \mathbf{0}$$
$$\mathbf{A} : \boldsymbol{\theta}_2 \neq \mathbf{0}.$$

The averaged error probabilities, α_{φ^*} and β_{φ^*}, were calculated using the Monte Carlo method with values generated from the following distributions:

- **Model (11) under H**

$$\boldsymbol{\theta}_1^{(j)} = 0$$

$$\sigma^{2(j)} | \boldsymbol{\theta}_1^{(j)} = 0 \sim IG\left(a_0 + \frac{1}{2}, b_0 + \frac{(\boldsymbol{\theta}_1^{(j)} - \mathrm{m}_0)^\top \mathrm{V}_0^{-1}(\boldsymbol{\theta}_1^{(j)} - \mathrm{m}_0)}{2} \right)$$

$$\mathbf{Y}^{(j)} | \sigma^{2(j)}, \boldsymbol{\theta}_1^{(j)} \sim N_n(\mathbf{1}_n \boldsymbol{\theta}_1^{(j)}, \sigma^{2(j)} \mathbf{I}_n).$$

- **Model (11) under A**

$$\sigma^{2(j)} \sim IG(a_0, b_0)$$
$$\boldsymbol{\theta}_1^{(j)} | \sigma^{2(j)} \sim N(\mathrm{m}_0, \sigma^{2(j)} \mathrm{V}_0)$$
$$\mathbf{Y}^{(j)} | \sigma^{2(j)}, \boldsymbol{\theta}_1^{(j)} \sim N_n(\mathbf{1}_n \boldsymbol{\theta}_1^{(j)}, \sigma^{2(j)} \mathbf{I}_n).$$

- **Model (12) under H**

$$\theta_2^{(j)} = \mathbf{0}$$

$$\theta_1^{(j)}|\theta_2^{(j)} = \mathbf{0} \sim t_s\left(2a_0 + 1; \mathbf{m_{01.2}}(\theta_2^{(j)}), \frac{2b_0 + (\theta_2^{(j)} - \mathbf{m_{02}})^\top \mathbf{V_{022}^{-1}}(\theta_2^{(j)} - \mathbf{m_{02}})}{2a_0 + 1}\mathbf{V_{011.2}}\right)$$

$$\sigma^{2(j)}|\theta_1^{(j)}, \theta_2^{(j)} = \mathbf{0} \sim IG\left(a_0 + 1, b_0 + \frac{(\theta^{(j)} - \mathbf{m_0})^\top \mathbf{V_0}^{-1}(\theta^{(j)} - \mathbf{m_0})}{2}\right)$$

$$\mathbf{Y}^{(j)}|\sigma^{2(j)}, \theta_1^{(j)}, \theta_2^{(j)} = \mathbf{0} \sim N_n(\mathbf{X}\theta^{(j)}, \sigma^{2(j)}\mathbf{I}_n).$$

- **Model (12) under A**

$$\sigma^{2(j)} \sim IG(a_0, b_0)$$

$$\theta^{(j)}|\sigma^{2(j)} \sim N_p(\mathbf{m_0}, \sigma^{2(j)}\mathbf{V_0})$$

$$\mathbf{Y}^{(j)}|\sigma^{2(j)}, \theta^{(j)} \sim N_n(\mathbf{X}\theta^{(j)}, \sigma^{2(j)}\mathbf{I}_n).$$

Then, $\mathbf{y}^{(j)} = (y_1^{(j)}, \ldots y_n^{(j)})$ is a random sample of the conditional distribution of $\mathbf{Y}$, $j = 1 \ldots M$.

In a first stage, we considered model (11) where $\theta = \theta_1$ and model (12) with $\theta = (\theta_1, \theta_2)^\top$. Note that the dimensionality of the parameter space, denoted by d, is different in the two models: for model (11), the dimensionality is $d = 2$ and for model (12), the dimensionality is $d = 3$. Samples of size $M = 1000$ were generated for each model under the respective hypotheses and also for different sample sizes between $n = 10$ and $n = 5000$. In model (12), the covariate x_{i1}, $i = 1 \ldots n$, was generated from a standard normal distribution. Finally, to obtain the adaptive values α_{φ^*} and β_{φ^*}, the two types of errors were considered as equally severe, that is, $a = b = 1$.

Figure 1 shows the averaged error probabilities for the FBST as functions of k for a sample size $n = 100$. This was replicated for all sample sizes in order to numerically find the corresponding k^* value that minimizes $\alpha_{\varphi_e} + \beta_{\varphi_e}$. Tables 1 and 2 and Figures 2 and 3 present the k^* and $\alpha_{\varphi_P^*}$ values as function of n for each model. As can be seen, both values have a decreasing trend when the sample size increases. In the case of the cutoff value for the evidence, it is possible to notice the differences in the results when the dimensionality of the parameter space change. Then, the k^* value depends not only on the sample size but also on the dimensionality of the parameter space, more specifically, it is greater when d is higher. However, this does not occur with $\alpha_{\varphi_P^*}$, which maintains almost the same values even if d increases. On the other hand, Figures 4 and 5 illustrate that in all these models, the optimal averaged error probabilities and their linear combination also decrease with increasing sample size.

Table 1. Cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n, with $d = 2$ and $d = 3$.

	k^*	
n	$d = 2$	$d = 3$
10	0.32530	0.51220
50	0.12534	0.22442
100	0.11705	0.21081
150	0.10889	0.19735
200	0.10092	0.18416
250	0.09323	0.17132
300	0.08587	0.15894
350	0.07893	0.14713
400	0.07243	0.13598
450	0.06641	0.12560
500	0.06091	0.11606
1000	0.03035	0.06689
1500	0.02223	0.07086
2000	0.01892	0.07173

(a) $y_i = \theta_1 + \varepsilon_i$, $\mathbf{H} : \theta_1 = 0$,

$m_0 = 0$, $V_0 = 1$, $a_0 = 2.01$, $b_0 = 1.01$,

$d = 2$.

(b) $y_i = \theta_1 + \theta_2 \, x_{i1} + \varepsilon_i$, $\mathbf{H} : \theta_2 = 0$,

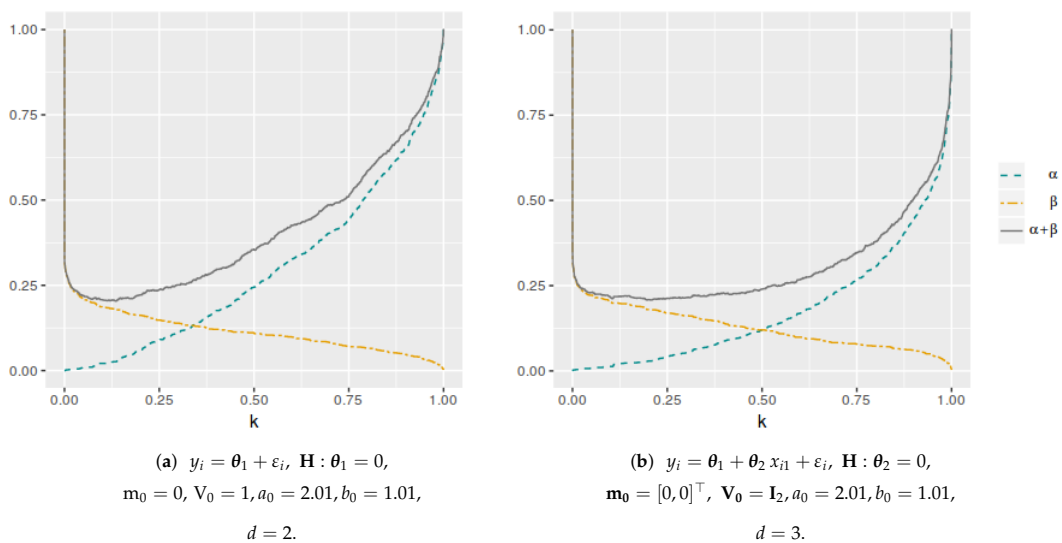$\mathbf{m_0} = [0,0]^\top$, $\mathbf{V_0} = \mathbf{I}_2$, $a_0 = 2.01$, $b_0 = 1.01$,

$d = 3$.

Figure 1. Averaged error probabilities (α_{φ_e}, β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k. Sample size $n = 100$.
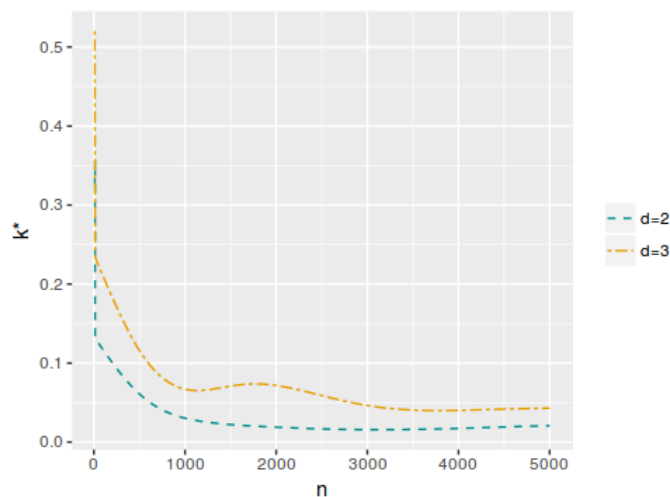
Figure 2. Cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n, with $d = 2$ and $d = 3$.

Table 2. Optimal averaged type-I error probability (α_{φ^*}) as a function of n, with $d = 2$ and $d = 3$.

	α_{φ^*}	
n	$d = 2$	$d = 3$
10	0.12400	0.09200
50	0.04515	0.04327
100	0.03899	0.03775
150	0.03327	0.03252
200	0.02817	0.02772
250	0.02380	0.02341
300	0.02018	0.01963
350	0.01732	0.01642
400	0.01513	0.01376
450	0.01353	0.01163
500	0.01241	0.01002
1000	0.00941	0.00683
1500	0.00827	0.00398
2000	0.00681	0.00524
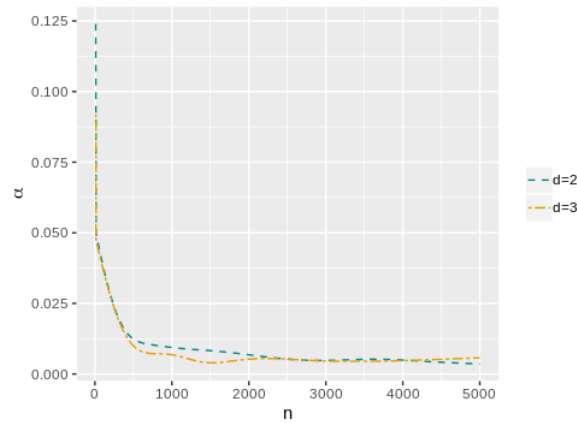
Figure 3. Optimal averaged type-I error probability (α_{φ^*}) as a function of n, with $d = 2$ and $d = 3$.

(**a**) $y_i = \theta_1 + \varepsilon_i$, $\mathbf{H} : \theta_1 = 0$,
$\mathrm{m}_0 = 0, \mathrm{V}_0 = 1, a_0 = 2.01, b_0 = 1.01$,
$d = 2$.

(**b**) $y_i = \theta_1 + \theta_2\, x_{i1} + \varepsilon_i$, $\mathbf{H} : \theta_2 = 0$,
$\mathbf{m_0} = [0,0]^\top$, $\mathbf{V_0} = \mathbf{I}_2, a_0 = 2.01, b_0 = 1.01$,
$d = 3$.

Figure 4. Unknown-variance model optimal averaged error probabilities ($\alpha^*_{\varphi^*_e}$, $\beta^*_{\varphi^*_e}$ and $\alpha^*_{\varphi^*_e} + \beta^*_{\varphi^*_e}$) as functions of n.
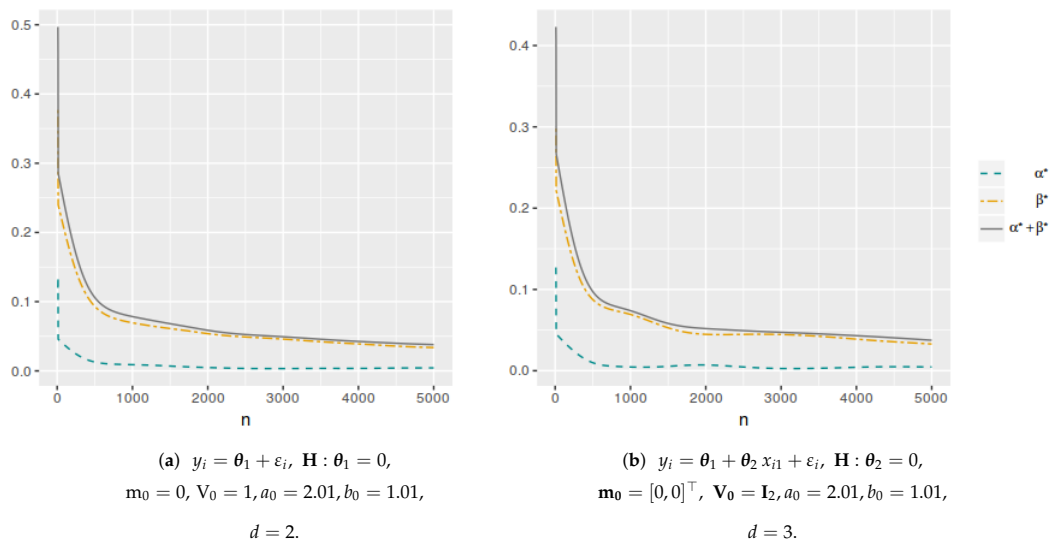

(**a**) $y_i = \theta_1 + \varepsilon_i$, $\mathbf{H} : \theta_1 = 0$,
$\mathrm{m}_0 = 0, \mathrm{V}_0 = 1, a_0 = 2.01, b_0 = 1.01$,
$d = 2$.

(**b**) $y_i = \theta_1 + \theta_2\, x_{i1} + \varepsilon_i$, $\mathbf{H} : \theta_2 = 0$,
$\mathbf{m_0} = [0,0]^\top$, $\mathbf{V_0} = \mathbf{I}_2, a_0 = 2.01, b_0 = 1.01$,
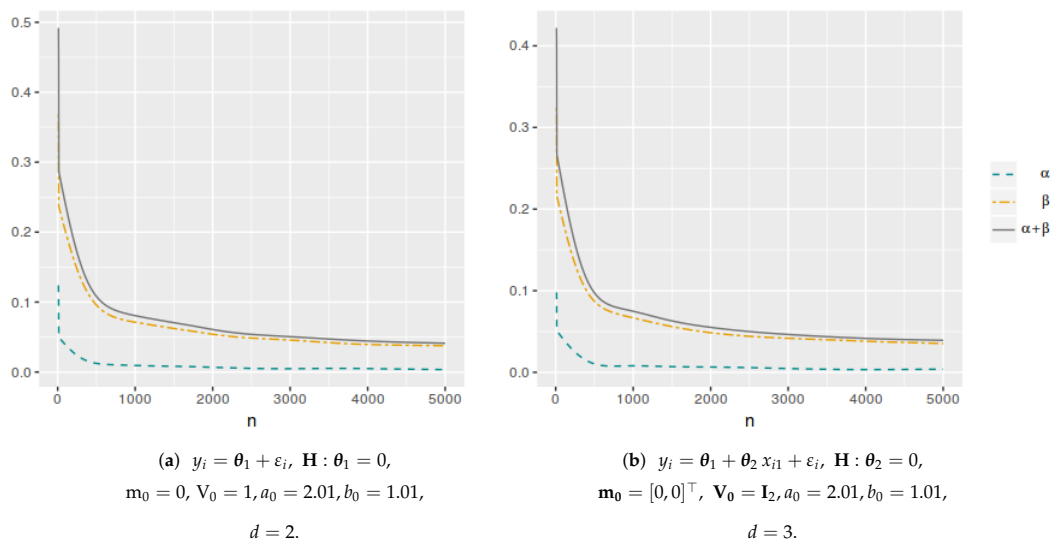$d = 3$.

Figure 5. Optimal averaged error probabilities (α_{φ^*}, β_{φ^*} and $\alpha_{\varphi^*} + \beta_{\varphi^*}$) as functions of n.

We choose a single random sample $\mathbf{y}_0$ to calculate the *e*-value and *P*-value for the models. Table 3 displays the results: the cases where **H** is rejected being represented by the

cells in **boldface**. It can be observed that the decision remains the same regardless of the index used.

Table 3. Cutoff values k^*, $ev(\mathbf{H}; \mathbf{y}_0)$ and P-value $(\mathbf{y}_0)$ as function of n, with $d = 2$ and $d = 3$.

	$d = 2$				$d = 3$			
n	k^*	ev	$\alpha_{\varphi_P^*}$	Pv	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
10	0.3253	0.9838	0.1240	0.7510	0.5122	0.9696	0.0920	0.4850
50	0.1253	**0.0820**	0.0451	**0.0190**	0.2244	0.9261	0.0433	0.3570
100	0.1171	**0.0000**	0.0390	**0.0000**	0.2108	0.4176	0.0377	0.0650
150	0.1089	**0.0973**	0.0333	**0.0200**	0.1974	0.2965	0.0325	0.0510
200	0.1009	**0.0036**	0.0282	**0.0000**	0.1842	**0.0466**	0.0277	**0.0040**
250	0.0932	**0.0001**	0.0238	**0.0000**	0.1713	**0.0620**	0.0234	**0.0050**
300	0.0859	**0.0000**	0.0202	**0.0000**	0.1589	**0.0119**	0.0196	**0.0010**
350	0.0789	**0.0000**	0.0173	**0.0000**	0.1471	**0.0282**	0.0164	**0.0030**
400	0.0724	**0.0000**	0.0151	**0.0000**	0.1360	**0.0347**	0.0138	**0.0020**
450	0.0664	**0.0000**	0.0135	**0.0000**	0.1256	**0.0628**	0.0116	**0.0040**
500	0.0609	**0.0000**	0.0124	**0.0000**	0.1161	**0.0181**	0.0100	**0.0010**
1000	0.0303	**0.0000**	0.0094	**0.0000**	0.0669	**0.0000**	0.0068	**0.0010**
1500	0.0222	**0.0000**	0.0083	**0.0000**	0.0709	**0.0000**	0.0040	**0.0010**
2000	0.0189	**0.0000**	0.0068	**0.0000**	0.0717	**0.0000**	0.0052	**0.0010**

As the second stage in our simulation study, we set two sample sizes $n = 60$ and $n = 120$ to perform the tests for model (12), increasing the dimensionality of the parameter space. In that scenario, the vector of coefficients was such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2)^\top$ and the hypotheses to be tested were

$$\mathbf{H} : \boldsymbol{\theta}_2 = 0$$
$$\mathbf{A} : \boldsymbol{\theta}_2 \neq 0.$$

So, by varying the dimension of vector $\boldsymbol{\theta}_1$, the different models considered for each test were obtained. Tables 4 and 5 and Figures 6 and 7 show the k^* and $\alpha_{\varphi_P^*}$ values as functions of d. For $d = 2$, the values correspond to model (11). We can say that, for a fixed hypothesis, the larger the dimensionality of the parameter space, the greater the value of k^*. In the case of the $\alpha_{\varphi_P^*}$ value, it does not change significantly when the dimensionality of the parameter space increases, except when the number of parameters is very large in relation to the sample size.

Table 4. Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d, with $n = 60$ and $n = 120$.

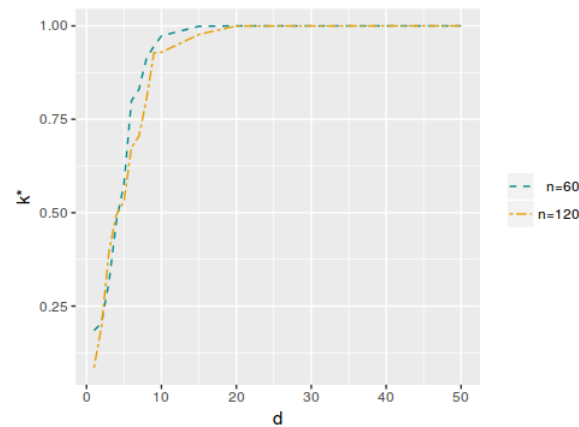|	k^*	
d	$n = 60$	$n = 120$
2	0.18500	0.08560
3	0.20420	0.19480
4	0.31510	0.39630
5	0.47790	0.49500
6	0.57670	0.53040
7	0.79970	0.67400
8	0.82970	0.70490
9	0.91250	0.80310
10	0.94540	0.92770
11	0.97300	0.92940
21	0.99990	0.99960
31	0.99990	0.99970
41	0.99990	0.99990
51	0.99990	0.99990

Figure 6. Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d, with $n = 60$ and $n = 120$.

Table 5. Optimal averaged type-I error probability (α_{φ^*}) as a function of d, with $n = 60$ and $n = 120$.

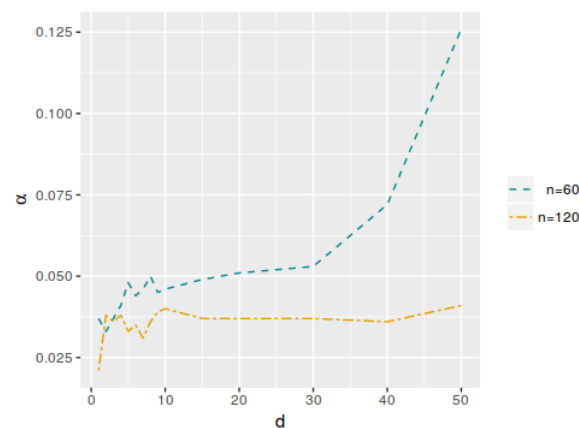|	α_{φ^*}	
d	$n = 60$	$n = 120$
2	0.03700	0.02100
3	0.03300	0.03800
4	0.03700	0.03600
5	0.04100	0.03800
6	0.04800	0.03300
7	0.04400	0.03500
8	0.04600	0.03100
9	0.05000	0.03600
10	0.04500	0.03900
11	0.04600	0.04000
21	0.05100	0.03700
31	0.05300	0.03700
41	0.07200	0.03600
51	0.12600	0.04100

Figure 7. Optimal averaged type-I error probability (α_{φ^*}) as a function of d, with $n = 60$ and $n = 120$.

Table 6 presents the e-value and P-value calculated for a single random sample $\mathbf{y}_0$. Here, with the e-value the null hypothesis is less easily rejected. This may be related to two things: it may be due to approximation error as a result of the simulation process or due to the fact that the evidence apparently converges to 1 as the dimensionality of the parameter space increases, in which case a more detailed study is required.

Table 6. Cutoff values k^*, $ev(\mathbf{H}; \mathbf{y}_0)$ and P-value $(\mathbf{y}_0)$ as functions of d, with $n = 60$ and $n = 120$.

	$n = 60$				$n = 120$			
d	k^*	ev	$\alpha_{\varphi_P^*}$	Pv	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
2	0.1850	0.6865	0.0370	0.3660	0.0856	**0.0082**	0.0210	**0.0010**
3	0.2042	0.5849	0.0330	0.1360	0.1948	0.7199	0.0380	0.1760
4	0.3151	0.8119	0.0370	0.1820	0.3963	0.9230	0.0360	0.2470
5	0.4779	**0.0000**	0.0410	**0.0000**	0.4950	**0.0000**	0.0380	**0.0010**
6	0.5767	**0.5672**	0.0480	**0.0290**	0.5304	0.7002	0.0330	0.0360
7	0.7997	0.8854	0.0440	0.0820	0.6740	0.9992	0.0350	0.2860
8	0.8297	**0.3267**	0.0460	**0.0050**	0.7049	0.7858	0.0310	**0.0260**
9	0.9125	**0.1919**	0.0500	**0.0020**	0.8031	**0.0009**	0.0360	**0.0010**
10	0.9454	**0.0006**	0.0450	**0.0010**	0.9277	**0.0001**	0.0390	**0.0010**
11	0.9730	**0.0000**	0.0460	**0.0000**	0.9294	**0.0000**	0.0400	**0.0000**
21	0.9999	**0.0000**	0.0510	**0.0000**	0.9996	**0.0000**	0.0370	**0.0000**
31	0.9999	1.0000	0.0530	**0.0240**	0.9997	**0.0495**	0.0370	**0.0010**
41	0.9999	**0.9998**	0.0720	**0.0010**	0.9999	**0.0004**	0.0360	**0.0010**
51	0.9999	1.0000	0.1260	**0.0000**	0.9999	**0.0000**	0.0410	**0.0000**

5. Numerical Examples

In this section, we present two applications with real datasets. We choose $a_0 = 3$ and $b_0 = 2$ as parameters of the inverse gamma prior distribution for σ^2. Additionally, in the normal prior for $\boldsymbol{\theta}$ given σ^2, $\mathbf{m}_0 = \mathbf{0}_{p \times 1}$ and $\mathbf{V}_0 = \mathbf{I}_p$ are taken as parameters. The Monte Carlo approximations were made generating samples of size $M = 10{,}000$.

5.1. Budget Shares of British Households Dataset

We select a dataset that draws 1519 observations from the 1980–1982 British Family Expenditure Surveys (FES) ([25]). In our application, we want to fit the model

$$y_i = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2\, x_{i1} + \boldsymbol{\theta}_3\, x_{i2} + \boldsymbol{\theta}_4\, x_{i3} + \boldsymbol{\theta}_5\, x_{i4} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2). \tag{13}$$

We consider as explanatory variables, respectively, the total net household income (rounded to the nearest 10 UK pounds sterling) (x_1), the budget share for alcohol expenditure (x_2), the budget share for fuel expenditure, and the age of household head (x_3). We take the budget share for food expenditure as the dependent variable (y). All the expenditures and income are measured in pounds sterling per week.

Table 7 summarizes the results for the hypotheses $\mathbf{H} : \boldsymbol{\theta}_j = 0$, $j = 1 \ldots 5$, by performing the test with the *p*-value at 0.05 significance level and also the *e*-value and the *P*-value with their respective adaptive significance levels. The cases where $\mathbf{H}$ is rejected are represented by the cells in **boldface**. $\hat{\boldsymbol{\theta}}_{Freq}$ and $\hat{\boldsymbol{\theta}}_{Bayes}$ are, respectively, the classical maximum likelihood estimator and the Bayes estimator of $\boldsymbol{\theta}$. It can be seen that unlike the *p*-value, the *e*-value and the *P*-value do not reject the hypothesis of nullity of the coefficient associated with the age of household head variable.

Table 7. Budget shares of British households dataset hypothesis-testing summary.

Coefficients	$\hat{\theta}_{Freq}$	α	pv	$\hat{\theta}_{Bayes}$	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
Intercept	0.3758	0.0500	**0.0000**	0.3700	0.7078	**0.0000**	0.0382	**0.0000**
x_{i1}	-0.0004	0.0500	**0.0000**	-0.0004	0.0113	**0.0000**	0.0001	**0.0000**
x_{i2}	-0.1533	0.0500	**0.0003**	-0.1283	0.9410	**0.1890**	0.1278	**0.0172**
x_{i3}	0.1717	0.0500	**0.0007**	0.1487	0.9520	**0.1957**	0.1468	**0.0143**
x_{i4}	0.0009	0.0500	**0.0119**	0.0010	0.0764	0.3048	0.0004	0.0666

Table 8 exposes the optimal averaged error probabilities using the *e*-value and the *P*-value. It can be noted that the values are very similar with both methodologies.

Table 8. Budget shares of British households dataset optimal averaged error probabilities.

Coefficients	$\alpha^*_{\varphi^*_e}$	$\alpha_{\varphi^*_P}$	$\beta^*_{\varphi^*_e}$	$\beta_{\varphi^*_P}$
Intercept	0.0466	0.0382	0.2157	0.2193
x_{i1}	0.0000	0.0001	0.0006	0.0006
x_{i2}	0.1521	0.1278	0.4146	0.4145
x_{i3}	0.1508	0.1468	0.4679	0.4410
x_{i4}	0.0004	0.0004	0.0080	0.0083

5.2. Boston Housing Dataset

We also take a dataset that contains information about housing values obtained from census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970 ([26]). These data are composed of 506 samples and 14 variables. The regression model we use is

$$y_i = \theta_1 + \theta_2\, x_{i1} + \theta_3\, x_{i2} + \theta_4\, x_{i3} + \theta_5\, x_{i4} + \theta_6\, x_{i5} + \theta_7\, x_{i6} + \theta_8\, x_{i7} + \theta_9\, x_{i8} + \theta_{10}\, x_{i9} + \varepsilon_i, \quad (14)$$

$$\varepsilon_i \sim N(0, \sigma^2).$$

We choose the following explanatory variables to fit our model: per capita crime rate by town (x_1), the proportion of residential land zoned for lots over 25.000 sq. ft (x_2), the proportion of non-retail business acres per town (x_3), the proportion of non-retail business acres per town (x_4), the average number of rooms per dwelling (x_5), the proportion of owner-occupied units built prior to 1940 (x_6), the weighted mean of distances to five Boston employment centers (x_7), the full-value property tax rate per 10.000 (x_8), the pupil–teacher ratio by town, and $1000(Bk - 0.63)^2$, where Bk is the proportion of black people by town (x_9). The dependent variable is the median value of the owner-occupied homes (in 1000 s) in the census tract (y).

The results for the hypotheses $\mathbf{H}: \theta_j = 0$, $j = 1 \dots 10$ by performing the test with the *p*-value, the *e*-value and the *P*-value, are summarized in Table 9. In this case, with the *e*-value the null hypotheses are less rejected. The *e*-value does not reject the hypotheses of nullity of the coefficients associated with the proportion of residential land zoned for lots over 25.000 sq. ft and proportion of non-retail business acres per town variables, while the *p*-value does. On the other hand, the *P*-value, unlike the *p*-value, does not reject the hypothesis for the proportion of residential land zoned for lots over 25.000 sq. ft variable, but it does for the Intercept. As can be observed in Table 10, for these data, the optimal averaged error probabilities values are also very close.

Table 9. Boston housing dataset hypothesis-testing summary.

Coefficients	$\hat{\theta}_{Freq}$	α	pv	$\hat{\theta}_{Bayes}$	k^*	ev	$\alpha_{\varphi^*_P}$	Pv
Intercept	1.7035	0.0500	0.6958	1.2035	0.9998	1.0000	0.1916	**0.0085**
x_{i1}	−0.1244	0.0500	**0.0006**	−0.1244	0.5780	**0.3365**	0.0010	**0.0001**
x_{i2}	0.0359	0.0500	**0.0224**	0.0362	0.4089	0.9012	0.0004	0.0025
x_{i3}	−0.1489	0.0500	**0.0235**	−0.1473	0.6390	0.9114	0.0025	**0.0023**
x_{i4}	6.7165	0.0500	**0.0000**	6.7336	0.9296	**0.0000**	0.0143	**0.0000**
x_{i5}	−0.0655	0.0500	**0.0000**	−0.0648	0.3275	**0.0141**	0.0001	**0.0000**
x_{i6}	−1.3198	0.0500	**0.0000**	−1.3091	0.8146	**0.0001**	0.0095	**0.0000**
x_{i7}	−0.0030	0.0500	0.2324	−0.0030	0.0124	0.9996	0.0002	0.0198
x_{i8}	−0.7652	0.0500	**0.0000**	−0.7528	0.8223	**0.0003**	0.0053	**0.0000**
x_{i9}	0.0145	0.0500	**0.0000**	0.0147	0.0297	**0.0113**	0.0001	**0.0000**

Table 10. Boston housing dataset optimal averaged error probabilities.

Coefficients	$\alpha^*_{\varphi^*_e}$	$\alpha_{\varphi^*_P}$	$\beta^*_{\varphi^*_e}$	$\beta_{\varphi^*_P}$
Intercept	0.1321	0.1916	0.6494	0.4946
x_{i1}	0.0018	0.0010	0.0165	0.0173
x_{i2}	0.0006	0.0004	0.0075	0.0079
x_{i3}	0.0030	0.0025	0.0286	0.0292
x_{i4}	0.0222	0.0143	0.1123	0.1181
x_{i5}	0.0000	0.0001	0.0068	0.0068
x_{i6}	0.0091	0.0095	0.0825	0.0808
x_{i7}	0.0000	0.0002	0.0016	0.0015
x_{i8}	0.0081	0.0053	0.0494	0.0521
x_{i9}	0.0000	0.0001	0.0019	0.0017

6. Conclusions

In this work, we present a method to find a cutoff value k^* for the Bayesian evidence in the FBST by minimizing the linear combination of the averaged type-I and type-II error probabilities for a given sample size n and also for a given dimensionality d of the parameter space in the context of linear regression models with unknown variance under the Bayesian perspective. In that sense, we provide a solution to the existing problem in the usual approach of hypothesis-testing procedures based on fixed cutoffs for measures of evidence: the increase of the sample size leads to the rejection of the null hypothesis. Furthermore, we compare our results with those obtained by using the test proposed by the authors of [13,14]. With our suggestion of cutoff value for the evidence in the FBST and also with the procedure proposed by the authors of [13,14], increasing the sample size implies that the probabilities of both kinds of optimal averaged errors and their linear combination decrease, unlike most cases, where, by setting a single level of significance independent of sample size, only type-II error probability decreases.

A detailed study is still needed for more complex models, so the methodology we propose to determine the adaptive cutoff value for evidence in the FBST could be extended to models with different prior specifications, which would involve, among other things, using approximate methods to find the prior predictive densities under the null and alternative hypotheses.

Author Contributions: Conceptualization, A.E.P.H., V.F., L.G.E. and C.A.d.B.P.; Methodology, A.E.P.H., V.F., L.G.E. and C.A.d.B.P.; Formal analysis, A.E.P.H., V.F. and L.G.E.; Investigation, A.E.P.H. and V.F.; Writing—review & editing, A.E.P.H. All authors have read and agreed to the published version of the manuscript.

Appendix A

As stated in Section 2, the normal linear regression model in (1) shows that the conditional distribution of **y** given parameters $(\boldsymbol{\theta}, \sigma^2)$ is the multivariate normal distribution $N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n)$. Therefore, the likelihood becomes

$$f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \right\}. \tag{A1}$$

The natural conjugate prior distribution of $(\boldsymbol{\theta}, \sigma^2)$ is a p-variate normal-inverse gamma distribution with hyperparameters $\mathbf{m_0}$, $\mathbf{V_0}$, a_0, and b_0, denoted by $(\boldsymbol{\theta}, \sigma^2) \sim N_p IG(\mathbf{m_0}, \mathbf{V_0}, a_0, b_0)$ ([20–22]):

$$g(\boldsymbol{\theta}, \sigma^2) = \frac{(b_0)^{a_0}}{(2\pi)^{p/2} |\mathbf{V_0}|^{1/2} \Gamma(a_0)} (\sigma^2)^{-\left(a_0 + \frac{p}{2} + 1\right)} \exp\left\{ -\frac{1}{2\sigma^2} \left[(\boldsymbol{\theta} - \mathbf{m_0})^\top \mathbf{V_0}^{-1} (\boldsymbol{\theta} - \mathbf{m_0}) + 2b_0 \right] \right\}, \tag{A2}$$

such that the conditional prior distributions of $\boldsymbol{\theta}$ given σ^2 is

$$g(\boldsymbol{\theta}|\sigma^2) = (2\pi)^{-p/2} |\mathbf{V_0}|^{-1/2} (\sigma^2)^{-p/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[(\boldsymbol{\theta} - \mathbf{m_0})^\top \mathbf{V_0}^{-1} (\boldsymbol{\theta} - \mathbf{m_0}) \right] \right\}, \tag{A3}$$

and the prior marginal distribution of σ^2 is

$$g(\sigma^2) = \frac{(b_0)^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-(a_0+1)} \exp\left\{ -\frac{b_0}{\sigma^2} \right\}, \tag{A4}$$

denoted, respectively, by

$$\boldsymbol{\theta}|\sigma^2 \sim N_p(\mathbf{m_0}, \sigma^2 \mathbf{V_0}), \quad \sigma^2 \sim IG(a_0, b_0). \tag{A5}$$

Both distributions are equivalent to the following new pair of distributions

$$g(\sigma^2|\boldsymbol{\theta}) = \frac{\left(b_0 + \frac{(\boldsymbol{\theta} - \mathbf{m_0})^\top \mathbf{V_0}^{-1} (\boldsymbol{\theta} - \mathbf{m_0})}{2} \right)^{\left(a_0 + \frac{p}{2}\right)}}{\Gamma\left(a_0 + \frac{p}{2}\right)} (\sigma^2)^{-\left(a_0 + \frac{p}{2} + 1\right)} \times$$
$$\exp\left\{ -\frac{1}{2\sigma^2} \left[(\boldsymbol{\theta} - \mathbf{m_0})^\top \mathbf{V_0}^{-1} (\boldsymbol{\theta} - \mathbf{m_0}) + 2b_0 \right] \right\}, \tag{A6}$$

and

$$g(\boldsymbol{\theta}) = \frac{(2b_0)^{a_0} \Gamma\left(a_0 + \frac{p}{2}\right)}{\pi^{p/2} |\mathbf{V_0}|^{1/2} \Gamma(a_0)} \left\{ (\boldsymbol{\theta} - \mathbf{m_0})^\top \mathbf{V_0}^{-1} (\boldsymbol{\theta} - \mathbf{m_0}) + 2b_0 \right\}^{-\left(a_0 + \frac{p}{2}\right)}$$
$$\propto \left\{ 1 + (\boldsymbol{\theta} - \mathbf{m_0})^\top (2b_0 \mathbf{V_0})^{-1} (\boldsymbol{\theta} - \mathbf{m_0}) \right\}^{-\left(a_0 + \frac{p}{2}\right)}. \tag{A7}$$

The density in (A7) is a p-variate t distribution with $2a_0$ degrees of freedom and hyperparameters $\mathbf{m_0}$ and $(b_0/a_0)\mathbf{V_0}$. Then, the distributions in (A6) and (A7) are denoted by

$$\sigma^2|\boldsymbol{\theta} \sim IG\left(a_0 + \frac{p}{2}, b_0 + \frac{(\boldsymbol{\theta} - \mathbf{m_0})^\top \mathbf{V_0}^{-1} (\boldsymbol{\theta} - \mathbf{m_0})}{2} \right), \quad \boldsymbol{\theta} \sim t_p\left(2a_0; \mathbf{m_0}, \frac{b_0}{a_0} \mathbf{V_0} \right). \tag{A8}$$

Now suppose that the $N_p IG(\mathbf{m_0}, \mathbf{V_0}, a_0, b_0)$ distribution (A2) is adopted as the prior distribution for $(\boldsymbol{\theta}, \sigma^2)$. Combining it with the likelihood (A1) gives the posterior distribution ([20–22]):

$$f(\boldsymbol{\theta}, \sigma^2|\mathbf{y}) \propto (\sigma^2)^{-\left(a_0 + \frac{n}{2} + \frac{p}{2} + 1\right)} \exp\left\{ -\frac{1}{2\sigma^2} \left[(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{V}^{*-1} (\boldsymbol{\theta} - \mathbf{m}^*) + 2b_1 \right] \right\}, \tag{A9}$$

where

$$\mathbf{V}^* = \left(\mathbf{V_0}^{-1} + \mathbf{X}^\top \mathbf{X} \right)^{-1}, \quad \mathbf{m}^* = \mathbf{V}^* \left(\mathbf{V_0}^{-1} \mathbf{m_0} + \mathbf{X}^\top \mathbf{y} \right),$$

$$a_1 = a_0 + \frac{n}{2}, \quad b_1 = b_0 + \frac{\mathbf{m_0}^\top \mathbf{V_0}^{-1} \mathbf{m_0} + \mathbf{y}^\top \mathbf{y} - \mathbf{m}^{*\top} \mathbf{V}^{*-1} \mathbf{m}^*}{2}.$$

If $\mathbf{X}^\top \mathbf{X}$ is non-singular, we can write

$$\mathbf{m}^* = \mathbf{V}^* \left(\mathbf{V_0}^{-1}\mathbf{m_0} + \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\theta}} \right),$$

where $\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$ is the classical maximum likelihood or least squares estimator of $\boldsymbol{\theta}$. Therefore, the posterior distribution of $(\boldsymbol{\theta}, \sigma^2)$ is

$$(\boldsymbol{\theta}, \sigma^2)|\mathbf{y} \sim N_p IG(\mathbf{m}^*, \mathbf{V}^*, a_1, b_1).$$

Consequently,

$$\boldsymbol{\theta}|\sigma^2, \mathbf{y} \sim N_p(\mathbf{m}^*, \sigma^2 \mathbf{V}^*), \quad \sigma^2|\mathbf{y} \sim IG(a_1, b_1), \tag{A10}$$

and this is equivalent to,

$$\sigma^2|\boldsymbol{\theta}, \mathbf{y} \quad \sim \quad IG\left(a_1 + \frac{p}{2}, b_1 + \frac{(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{V}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*)}{2}\right), \tag{A11}$$

$$\boldsymbol{\theta}|\mathbf{y} \quad \sim \quad t_p\left(2a_1; \mathbf{m}^*, \frac{b_1}{a_1}\mathbf{V}^*\right). \tag{A12}$$

Consider now conditional distributions given partial specification of $\boldsymbol{\theta}$. First let $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$, and consider distributions conditional on $\boldsymbol{\theta}_2$. Suppose that $(\boldsymbol{\theta}, \sigma^2) \sim N_p IG(\boldsymbol{m_0}, \mathbf{V_0}, a_0, b_0)$. Corresponding distributions result if we change a_0 to a_1, b_0 to b_1, $\mathbf{m_0}$ to $\mathbf{m}^*$ and $\mathbf{V_0}$ to $\mathbf{V}^*$. If $\boldsymbol{\theta}_1$ has s elements and $\boldsymbol{\theta}_2$ has r elements, write

$$\mathbf{m_0} = \begin{bmatrix} \mathbf{m_{01}} \\ \mathbf{m_{02}} \end{bmatrix}, \quad \mathbf{V_0} = \begin{bmatrix} \mathbf{V_{011}} & \mathbf{V_{012}} \\ \mathbf{V_{021}} & \mathbf{V_{022}} \end{bmatrix},$$

where $\mathbf{m_{01}}$ is $s \times 1$, $\mathbf{V_{011}}$ is $s \times s$, $\mathbf{m_{02}}$ is $r \times 1$, $\mathbf{V_{022}}$ is $r \times r$, with $r = p - s$. Now since $\boldsymbol{\theta}$ given σ^2 is distributed as $N_p(\mathbf{m_0}, \sigma^2 \mathbf{V_0})$, using general results on multivariate normal distributions (see [30]), we have the following distributions:

$$\boldsymbol{\theta}_2|\sigma^2 \quad \sim \quad N_r(\mathbf{m_{02}}, \sigma^2 \mathbf{V_{022}}), \tag{A13}$$

$$(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2, \sigma^2) \quad \sim \quad N_s(\mathbf{m_{01.2}}(\boldsymbol{\theta}_2), \sigma^2 \mathbf{V_{011.2}}), \tag{A14}$$

where $\mathbf{m_{01.2}}(\boldsymbol{\theta}_2) = \mathbf{m_{01}} + \mathbf{V_{012}}\mathbf{V_{022}}^{-1}(\boldsymbol{\theta}_2 - \mathbf{m_{02}})$ and $\mathbf{V_{011.2}} = \mathbf{V_{011}} - \mathbf{V_{012}}\mathbf{V_{022}}^{-1}\mathbf{V_{021}}$.

From (A13) and the prior distribution of σ^2 we have that

$$(\boldsymbol{\theta}_2, \sigma^2) \sim N_r IG(\mathbf{m_{02}}, \mathbf{V_{022}}, a_0, b_0) \tag{A15}$$

and hence

$$\boldsymbol{\theta}_2 \quad \sim \quad t_r\left(2a_0; \mathbf{m_{02}}, \frac{b_0}{a_0}\mathbf{V_{022}}\right), \tag{A16}$$

$$\sigma^2|\boldsymbol{\theta}_2 \quad \sim \quad IG\left(a_0 + \frac{r}{2}, b_0 + \frac{(\boldsymbol{\theta}_2 - \mathbf{m_{02}})^\top \mathbf{V_{022}}^{-1}(\boldsymbol{\theta}_2 - \mathbf{m_{02}})}{2}\right), \tag{A17}$$

Now (A14) and (A17) together give

$$(\boldsymbol{\theta}_1, \sigma^2|\boldsymbol{\theta}_2) \sim N_s IG\left(\mathbf{m_{01.2}}(\boldsymbol{\theta}_2), \mathbf{V_{011.2}}, a_0 + \frac{r}{2}, b_0 + \frac{(\boldsymbol{\theta}_2 - \mathbf{m_{02}})^\top \mathbf{V_{022}}^{-1}(\boldsymbol{\theta}_2 - \mathbf{m_{02}})}{2}\right) \tag{A18}$$

and finally

$$\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 \sim t_s \left(2a_0 + r; \mathbf{m}_{01.2}(\boldsymbol{\theta}_2), \frac{2b_0 + (\boldsymbol{\theta}_2 - \mathbf{m}_{02})^\top \mathbf{V}_{022}^{-1} (\boldsymbol{\theta}_2 - \mathbf{m}_{02})}{2a_0 + r} \mathbf{V}_{011.2} \right). \tag{A19}$$

Appendix B

Let $f_{\mathbf{H}}$ and $f_{\mathbf{A}}$ be the Bayesian prior predictive densities under the respective hypotheses **H** and **A** described in Section 3.1. Both are probability density functions over the sample space $\boldsymbol{\Omega}$, and they are calculated as the following conditional expectations:

$$f_{\mathbf{H}}(\mathbf{y}) = E_{\boldsymbol{\xi}}[f(\mathbf{y}|\boldsymbol{\xi})|\mathbf{H}]$$

$$= \int_{\mathbf{H}} f(\mathbf{y}|\boldsymbol{\xi}) \, dP_{\mathbf{H}}(\boldsymbol{\xi})$$

$$= \int_{\mathbf{H}} f(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \, g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \, d\boldsymbol{\theta}_1 \, d\boldsymbol{\theta}_2 \, d\sigma^2,$$

where $g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2)$ is the prior density under **H** calculated as

$$g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) = \frac{g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2)) \, \mathbb{1}(\boldsymbol{\theta}_2 = \mathbf{0})}{\oint_{\mathbf{H}} g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \, d\boldsymbol{\theta}_1 \, d\boldsymbol{\theta}_2 \, d\sigma^2}$$

$$= \frac{g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \, \mathbb{1}(\boldsymbol{\theta}_2 = \mathbf{0})}{\int_{\mathbb{R}^s \times \mathbb{R}_+} g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2) \, d\boldsymbol{\theta}_1 \, d\sigma^2}$$

$$= g(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}).$$

Thus, $f_{\mathbf{H}}(\mathbf{y})$ is given by

$$f_{\mathbf{H}}(\mathbf{y}) = \int_{\mathbf{H}} f(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \, g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \, d\boldsymbol{\theta}_1 \, d\boldsymbol{\theta}_2 \, d\sigma^2$$

$$= \int_{\mathbb{R}^s \times \mathbb{R}_+} f(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2) \, g(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}) \, d\boldsymbol{\theta}_1 \, d\sigma^2$$

$$= \int_{\mathbb{R}^s \times \mathbb{R}_+} N_n(\mathbf{X}\mathbf{C}\boldsymbol{\theta}_1, \sigma^2 \mathbf{I}_n) \times$$

$$N_s IG \left(\mathbf{m}_{01.2}(\mathbf{0}), \mathbf{V}_{011.2}, a_0 + \frac{r}{2}, b_0 + \frac{\mathbf{m}_{02}^\top (\mathbf{V}_{022})^{-1} \mathbf{m}_{02}}{2} \right) d\boldsymbol{\theta}_1 \, d\sigma^2$$

$$= t_n \left(2\left(a_0 + \frac{r}{2}\right); \mathbf{X}\mathbf{C}\mathbf{m}_{01.2}(\mathbf{0}), \frac{b_0 + \frac{\mathbf{m}_{02}^\top (\mathbf{V}_{022})^{-1} \mathbf{m}_{02}}{2}}{\left(a_0 + \frac{r}{2}\right)} \left(\mathbf{I}_n + (\mathbf{X}\mathbf{C}) \mathbf{V}_{011.2} (\mathbf{X}\mathbf{C})^\top \right) \right), \tag{A20}$$

where $\mathbf{C}_{(s+r) \times s} = [\mathbf{I}_s, \mathbf{0}_{s \times r}]^\top$.

The prior predictive density under **A** can be obtained as follows

$$f_{\mathbf{A}}(\mathbf{y}) = E_{\boldsymbol{\xi}}[\, f(\mathbf{y}|\boldsymbol{\xi})|\mathbf{A}\,]$$

$$= \int_{\mathbf{A}} f(\mathbf{y}|\boldsymbol{\xi})\, dP_{\mathbf{A}}(\boldsymbol{\xi})$$

$$= \int_{\mathbf{A}} f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2)\, g_{\mathbf{A}}(\boldsymbol{\theta}, \sigma^2)\, d\boldsymbol{\theta}\, d\sigma^2$$

$$= \int_{\mathbf{A}} f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2)\, g(\boldsymbol{\theta}, \sigma^2)\, d\boldsymbol{\theta}\, d\sigma^2$$

$$= \int_{\mathbf{A}} N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \times N_p IG(\mathbf{m_0}, \mathbf{V_0}, a_0, b_0)\, d\boldsymbol{\theta}\, d\sigma^2$$

$$= t_n\left(2a_0; \mathbf{X}\mathbf{m_0}, \frac{b_0}{a_0}\left(\mathbf{I}_n + \mathbf{X}\mathbf{V_0}\mathbf{X}^{\top}\right)\right). \tag{A21}$$

where $P_{\mathbf{H}}$ and $P_{\mathbf{A}}$ are the prior probability measure of $\boldsymbol{\xi}$ restricted to the sets $\mathbf{H}$ and $\mathbf{A}$, respectively.

References

1. Pereira, C.A.B.; Stern, J.M. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. *Entropy* **1999**, *1*, 99–110. [CrossRef]
2. Wagner, B.; Stern, J.M. The Rules of Logic Composition for the Bayesian Epistemic e-Values. *Logic J. IGPL* **2007**, *15*, 401–420.
3. Stern, J.M. Cognitive constructivism, eigen-solutions, and sharp statistical hypotheses. *Cybern. Hum. Knowing* **2007**, *14*, 9–46.
4. Madruga, R.; Esteves, L.G.; Wechsler, S. On the Bayesianity of Pereira-Stern tests. *TEST* **2001**, *10*, 291–299. [CrossRef]
5. Kempthorne, O.; Folks, L. *Probability, Statistics, and Data Analysis*; Iowa State University Press: Ames, IA, USA, 1971.
6. Cox, D.R.; Spjøtvoll, E.; Johansen, S.; van Zwet, W.R.; Bithell, J.F.; Barndorff-Nielsen, O.; Keuls, M. The Role of Significance Tests [with Discussion and Reply]. *Scand. J. Stat.* **1978**, *4*, 49–70.
7. Lindley, D.V.; Barndorff-Nielsen, O.; Gustav, E.; Harsaae, E.; Thorburn, D.; Hald, A.; Spjötvoll, E. The Bayesian Approach [with Discussion and Reply]. *Scand. J. Stat.* **1978**, *5*, 1–26.
8. Cox, D.R. Statistical significance tests. *Br. J. Clin. Pharmacol.* **1982**, *14*, 325–331. [CrossRef] [PubMed]
9. Johnstone, D.J.; Lindley, D.V. Bayesian inference given data 'significant at α': Tests of point hypotheses. *Theor. Decis.* **1995**, *38*, 51–60. [CrossRef]
10. Oliveira, M.C. Definition of the Level of Significance as a Function of the Sample Size. Masters Dissertation, Universidade de São Paulo, Instituto de Matemática e Estatística, Departamento de Estatística, São Paulo, Brazil, 2014. (In Portuguese)
11. Pereira, C.A.B. Test of Hypotheses Defined in Spaces of Different Dimensions: Bayesian View and Classical Interpretation. Ph.D. Thesis, Universidade de São Paulo, Instituto de Matemática e Estatstica, Departamento de Estatstica, São Paulo, Brazil, 1985. (In Portuguese)
12. DeGroot, M.H. *Probability and Statistics*, 2nd ed.; Addison-Whesley Publishing Company: Boston, MA, USA, 1986.
13. Pereira, C.A.B.; Nakano, E.Y.; Fossaluza, V.; Esteves, L.G.; Gannon, M.A.; Polpo, A. Hypothesis tests for bernoulli experiments: Ordering the sample space by bayes factors and using adaptive significance levels for decisions. *Entropy* **2017**, *19*, 696. [CrossRef]
14. Gannon, M.A.; Pereira, C.A.B.; Polpo, A. Blending bayesian and classical tools to define optimal sample-size-dependent significance levels. *Am. Stat.* **2019**, *73* (Suppl. S1), 213–222. [CrossRef]
15. Montoya-Delgado, L.E.; Irony, T.Z.; Pereira, C.A.B.; Whittle, M.R. An unconditional exact test for the Hardy-Weinberg equilibrium law: Sample-space ordering using the Bayes factor. *Genetics* **2001**, *158*, 875–883. [CrossRef] [PubMed]
16. Irony, T.Z.; Pereira, C.A.B. Bayesian hypothesis test: Using Surface Integrals to Distribute Prior Information among the Hypotheses. *Resenhas Inst. Matemática Estatística Univ. Paulo* **1995**, *2*, 27–46.
17. Pereira, C.A.B.; Wechsler, S. On the concept of P-value. *Braz. J. Probab. Stat.* **1993**, *7*, 159–177.
18. Esteves, L.G.; Izbicki, R.; Stern, J.M.; Stern, R.B. Pragmatic Hypotheses in the Evolution of Science. *Entropy* **2019**, *21*, 883. [CrossRef]
19. Schervish, M.J. *Theory of Statistics*; Springer: New York, NY, USA, 1995.
20. O'Hagan, A.; Forster, J.J. *Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference*; Arnold: London, UK, 2004.
21. Box, G.E.P.; Tiao, G.C. *Bayesian Inference in Statistical Analysis*; Wiley: New York, NY, USA, 1973.
22. DeGroot, M.H. *Optimal Statistical Decisions*; McGraw-Hill: New York, NY, USA, 1970.
23. Madruga, R.; Pereira, C.A.B. Power of FBST: Standard examples. *Inst. Interam. Estadística. Estadístic* **2005**, *57*, 1–9.
24. Berger, J.O.; Delampady, M. Pragmatic Testing Precise Hypotheses. *Stat. Sci.* **1987**, *22*, 317–335.
25. Blundell, R.; Duncan, A.; Pendakur, K. Semiparametric estimation and consumer demand. *J. Appl. Econom.* **1998**, *13*, 435–461. [CrossRef]
26. Harrison, D.; Rubinfeld, D.L. Hedonic Housing Prices and the Demand for Clean Air. *J. Environ. Econ. Manag.* **1978**, *5*, 81–102. [CrossRef]

27. Croissant, Y.; Graves, S. Ecdat: Data Sets for Econometrics. R Package Version 0.4-2. 2022. Available online: https://CRAN.R-project.org/package=Ecdat (accessed on 26 September 2022).

28. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*, 4th ed.; Springer: New York, NY, USA, 2002. Available online: https://www.stats.ox.ac.uk/pub/MASS4/ (accessed on 26 September 2022).

29. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022. Available online: https://www.R-project.org/ (accessed on 26 September 2022).

30. Mardia, K.V.; Kent, J.T.; Bibby, J.M. *Multivariate Analysis*; Academic Press: London, UK, 1979.