Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 2, 299–320
DOI: 10.1214/15-BJPS281
© Brazilian Statistical Association, 2016

Limiting behavior of the Jeffreys power-expected-posterior Bayes factor in Gaussian linear models

D. Fouskakis[a] **and I. Ntzoufras**[b]

[a]*National Technical University of Athens*
[b]*Athens University of Economics and Business*

Abstract. Expected-posterior priors (EPPs) have been proved to be extremely useful for testing hypotheses on the regression coefficients of normal linear models. One of the advantages of using EPPs is that impropriety of baseline priors causes no indeterminacy in the computation of Bayes factors. However, in regression problems, they are based on one or more *training samples*, that could influence the resulting posterior distribution. On the other hand, the *power-expected-posterior priors* are minimally-informative priors that reduce the effect of training samples on the EPP approach, by combining ideas from the power-prior and unit-information-prior methodologies. In this paper, we prove the consistency of the Bayes factors when using the power-expected-posterior priors, with the independence Jeffreys as a baseline prior, for normal linear models, under very mild conditions on the design matrix.

1 Introduction

Pérez and Berger (2002) developed priors for model comparison, through utilization of the device of "imaginary training samples" (Good, 2004; Spiegelhalter and Smith, 1988; Iwaki, 1997). They defined the expected-posterior prior (EPP) as the posterior distribution of a parameter vector for the model under consideration, averaged over all possible imaginary samples $\mathbf{y}^*$ coming from a "suitable" predictive distribution $m^*(\mathbf{y}^*)$. Hence, the EPP for the parameter vector $\boldsymbol{\theta}_\ell$, of any model $M_\ell \in \mathcal{M}$, with $\mathcal{M}$ denoting the model space, is

$$\pi_\ell^{EPP}(\boldsymbol{\theta}_\ell) = \int \pi_\ell^N(\boldsymbol{\theta}_\ell|\mathbf{y}^*)m^*(\mathbf{y}^*)\,d\mathbf{y}^*, \tag{1}$$

where $\pi_\ell^N(\boldsymbol{\theta}_\ell|\mathbf{y}^*)$ is the posterior of $\boldsymbol{\theta}_\ell$ for model M_ℓ using a baseline prior $\pi_\ell^N(\boldsymbol{\theta}_\ell)$ and data $\mathbf{y}^*$.

An attractive option for m^* arises from selecting a "reference" or "base" model M_0 for the training sample and defining $m^*(\mathbf{y}^*) = m_0^N(\mathbf{y}^*) \equiv f(\mathbf{y}^*|M_0)$ to be the prior predictive distribution, evaluated at $\mathbf{y}^*$, for the reference model M_0 under the baseline prior $\pi_0^N(\boldsymbol{\theta}_0)$. For the variable-selection problem considered in this

Key words and phrases. Bayesian variable selection, Bayes factors, consistency, expected-posterior priors, Gaussian linear models, objective model selection methods, power-expected-posterior priors, power prior, training sample, unit-information prior.

Received May 2014; accepted January 2015.

paper, the constant model (with no predictors) is used as a reference model, following the skeptical-prior approach described by Spiegelhalter, Abrams and Myles (2004, Section 5.5.2). This selection simplifies computations, and makes the EPP approach equivalent to the arithmetic intrinsic Bayes factor approach of Berger and Pericchi (1996).

One of the advantages of using EPPs is that impropriety of baseline priors causes no indeterminacy in the computation of Bayes factors. With EPPs, we can use an improper baseline prior $\pi_\ell^N(\boldsymbol{\theta}_\ell)$ in (1), since the arbitrary constants cancel out in the calculation of any Bayes factor. However, in regression problems, EPPs are based on one or more *training samples*, that could influence the resulting posterior distribution.

To diminish the effect of training samples on the EPP approach and simultaneously to produce a minimally-informative prior, Fouskakis, Ntzoufras and Draper (2015) introduced the *power-expected-posterior* (PEP) priors, by combining ideas from the power-prior approach of Ibrahim and Chen (2000) and the unit-information-prior approach of Kass and Wasserman (1995). As a first step, the likelihoods involved in the EPP distribution are raised to the power $1/\delta$ and then are density-normalized. This power parameter δ is set equal to the size of the training sample n^*, to represent information equal to one data point. Regarding the size of the training sample, n^*, this is set equal to the sample size n; in this way the selection of a training sample and its effect on the posterior model comparison is completely avoided.

In what follows, we examine variable-selection problems in Gaussian regression models. Thus, for any model M_ℓ, with parameters $\boldsymbol{\theta}_\ell = (\boldsymbol{\beta}_\ell, \sigma_\ell^2)$, the likelihood is specified by

$$(\mathbf{Y}|\mathrm{X}_\ell, \boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell) \sim N_n(\mathrm{X}_\ell \boldsymbol{\beta}_\ell, \sigma_\ell^2 \mathrm{I}_n), \tag{2}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is a vector containing the (real-valued) responses for all subjects, X_ℓ is a $n \times d_\ell$ design matrix containing the values of the explanatory variables in its columns, I_n is the $n \times n$ identity matrix, $\boldsymbol{\beta}_\ell$ is a vector of length d_ℓ summarizing the effects of the covariates in model M_ℓ on the response $\mathbf{Y}$ and σ_ℓ^2 is the error variance. Furthermore, we denote the imaginary/training data set by $\mathbf{y}^*$, their size by n^*, and the corresponding imaginary design matrix by X^* of size $n^* \times (p+1)$, where p denotes the total number of available covariates. Following the PEP methodology, we set $n^* = n$ and $\mathrm{X}^* = \mathrm{X}$, where X is the original $n \times (p+1)$ design matrix.

For any model $M_\ell \in \mathcal{M}$, we denote by $\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*)$ the baseline prior for model parameters $\boldsymbol{\beta}_\ell$ and σ_ℓ^2, with X_ℓ^* being the imaginary design matrix under model M_ℓ. Then the *power-expected-posterior* (PEP) prior, $\pi_\ell^{PEP}(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*, \delta)$, takes the following form:

$$\begin{aligned}\pi_\ell^{PEP}&(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*, \delta)\\ &= \pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*) \int \frac{m_0^N(\mathbf{y}^*|\mathrm{X}_0^*, \delta)}{m_\ell^N(\mathbf{y}^*|\mathrm{X}_\ell^*, \delta)} f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*, \delta)\, d\mathbf{y}^*,\end{aligned} \tag{3}$$

where $f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*, \delta) \propto f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*)^{1/\delta}$ is the likelihood, evaluated at $\mathbf{y}^*$, under model M_ℓ, raised to the power of $1/\delta$ and density-normalized, that is,

$$\begin{aligned} f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*, \delta) &= \frac{f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*)^{1/\delta}}{\int f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*)^{1/\delta}\, d\mathbf{y}^*} \\ &= \frac{f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_\ell^*\boldsymbol{\beta}_\ell, \sigma_\ell^2 \mathrm{I}_{n^*})^{1/\delta}}{\int f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_\ell^*\boldsymbol{\beta}_\ell, \sigma_\ell^2 \mathrm{I}_{n^*})^{1/\delta}\, d\mathbf{y}^*} \\ &= f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_\ell^*\boldsymbol{\beta}_\ell, \delta\sigma_\ell^2 \mathrm{I}_{n^*}); \end{aligned} \tag{4}$$

here $f_{N_d}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the density of the d-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, evaluated at $\mathbf{y}$.

When the reference model M_0 is nested in all other models (like in our case) the EPP (and therefore the PEP prior) for the parameter vector under M_0 is clearly the same as the baseline prior, that is,

$$\pi_0^{PEP}(\boldsymbol{\beta}_0, \sigma_0^2|\mathrm{X}_0^*, \delta) = \pi_0^N(\boldsymbol{\beta}_0, \sigma_0^2|\mathrm{X}_0^*),$$

with X_0^* being the imaginary design matrix under model M_0.

The distribution $m_\ell^N(\mathbf{y}^*|\mathrm{X}_\ell^*, \delta)$ appearing in (3) is the prior predictive distribution (or the marginal likelihood), evaluated at $\mathbf{y}^*$, of model M_ℓ, using the power likelihood defined in (4), under the baseline prior $\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*)$, that is,

$$m_\ell^N(\mathbf{y}^*|\mathrm{X}_\ell^*, \delta) = \iint f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_\ell^*\boldsymbol{\beta}_\ell, \delta\sigma_\ell^2 \mathrm{I}_{n^*}) \pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*)\, d\boldsymbol{\beta}_\ell\, d\sigma_\ell^2. \tag{5}$$

Similarly, the distribution $m_0^N(\mathbf{y}^*|\mathrm{X}_0^*, \delta)$ appearing in (3) is the prior predictive distribution, evaluated at $\mathbf{y}^*$, of the reference model M_0, using the power likelihood defined in (4) (with $\ell = 0$), under the baseline prior $\pi_0^N(\boldsymbol{\beta}_0, \sigma_0^2|\mathrm{X}_0^*)$, that is,

$$m_0^N(\mathbf{y}^*|\mathrm{X}_0^*, \delta) = \iint f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_0^*\boldsymbol{\beta}_0, \delta\sigma_0^2 \mathrm{I}_{n^*}) \pi_0^N(\boldsymbol{\beta}_0, \sigma_0^2|\mathrm{X}_0^*)\, d\boldsymbol{\beta}_0\, d\sigma_0^2. \tag{6}$$

Here, we use the independence Jeffreys prior (or reference prior) as the baseline prior distribution. Hence, for any $M_\ell \in \mathcal{M}$ we have

$$\pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma^2|\mathrm{X}_\ell^*) = \frac{c_\ell}{\sigma_\ell^2}, \tag{7}$$

where c_ℓ is an unknown normalizing constant; we refer to the resulting PEP prior as J-PEP.

It is worth noting that our method works in a totally different fashion than fractional Bayes factors (O'Hagan, 1995). In the latter, a fraction b of the full likelihood is used to "properize" the baseline prior and the remaining fraction $(1 - b)$ of the full likelihood is used for model comparison. In contrast, with our approach, the original likelihood is used only once, for simultaneous variable selection and posterior inference. Moreover, the fraction of the likelihood (power likelihood)—used in the expected-posterior expression of our prior distribution—refers solely to the imaginary data coming from a prior predictive distribution based on the reference model.

2 The conditional J-PEP prior distribution

In the following, under any model M_ℓ, we denote by

$$\mathrm{H}_\ell = \mathrm{X}_\ell (\mathrm{X}_\ell^T \mathrm{X}_\ell)^{-1} \mathrm{X}_\ell^T \quad \text{and by} \quad \mathrm{P}_\ell = \mathrm{I}_n - \mathrm{H}_\ell$$

and the corresponding measures based on X_ℓ^* by H_ℓ^* and P_ℓ^*, respectively.

Under (7), the corresponding marginal likelihood, with response data $\mathbf{y}^*$, design matrix X_ℓ^* and likelihood function raised to the power of $1/\delta$, is given by

$$m_\ell^N(\mathbf{y}^*|\mathrm{X}_\ell^*, \delta) = c_\ell \pi^{(d_\ell - n^*)/2} |\mathrm{X}_\ell^{T*} \mathrm{X}_\ell^*|^{-1/2} \Gamma\left(\frac{n^* - d_\ell}{2}\right) RSS_\ell^{*\,-(n^*-d_\ell)/2},$$

where RSS_ℓ^* is the residual sum of squares given by $RSS_\ell^* = \mathbf{y}^{*T} \mathrm{P}_\ell^* \mathbf{y}^*$. Similarly, in the rest of the paper we denote by $RSS_\ell = \mathbf{y}^T \mathrm{P}_\ell \mathbf{y}$.

The J-PEP prior for the parameters of model M_ℓ is given by

$$\begin{aligned}
&\pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \mathrm{X}_\ell^*, \delta) \\
&\quad = \int \pi_\ell^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \mathbf{y}^*; \mathrm{X}_\ell^*, \delta) m_0^N(\mathbf{y}^*|\mathrm{X}_0^*, \delta)\, d\mathbf{y}^* \\
&\quad = \int f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*, \delta) \pi^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*) \frac{m_0^N(\mathbf{y}^*|\mathrm{X}_0^*, \delta)}{m_\ell^N(\mathbf{y}^*|\mathrm{X}_\ell^*, \delta)}\, d\mathbf{y}^* \\
&\quad = \iint \Big[\int \big(f(\mathbf{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell; \mathrm{X}_\ell^*, \delta) f(\mathbf{y}^*|\boldsymbol{\beta}_0, \sigma_0^2, M_0; \mathrm{X}_0^*, \delta) \pi^N(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\mathrm{X}_\ell^*)\big) \\
&\qquad\qquad / (m_\ell^N(\mathbf{y}^*|\mathrm{X}_\ell^*, \delta))\, d\mathbf{y}^* \Big] \\
&\qquad \times \pi_0^N(\boldsymbol{\beta}_0, \sigma_0^2|\mathrm{X}_0^*)\, d\boldsymbol{\beta}_0\, d\sigma_0^2 \\
&\quad = \iint \pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{\beta}_0, \sigma_0^2; \mathrm{X}_\ell^*, \delta) \pi_0^N(\boldsymbol{\beta}_0, \sigma_0^2|\mathrm{X}_0^*)\, d\boldsymbol{\beta}_0\, d\sigma_0^2
\end{aligned}$$

with the conditional J-PEP prior given by

$$\begin{aligned}
&\pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{\beta}_0, \sigma_0^2; \mathrm{X}_\ell^*, \delta) \\
&\quad = \int \frac{f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_\ell \boldsymbol{\beta}_\ell, \delta\sigma_\ell^2 \mathrm{I}_{n^*}) f_{N_{n^*}}(\mathbf{y}^*; \mathrm{X}_0 \boldsymbol{\beta}_0, \delta\sigma_0^2 \mathrm{I}_{n^*}) c_\ell / \sigma_\ell^2}{c_\ell \pi^{(d_\ell - n^*)/2} |\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*|^{-1/2} \Gamma((n^* - d_\ell)/2) RSS_\ell^{*\,-(n^*-d_\ell)/2}}\, d\mathbf{y}^*
\end{aligned}$$

$$= \frac{\pi^{-(d_\ell - n^*)/2}}{\sigma_\ell^2 \Gamma((n^* - d_\ell)/2)} |X_\ell^{*T} X_\ell^*|^{1/2} \times \int RSS_\ell^{*(n^*-d_\ell)/2} f_{N_{n^*}}(\mathbf{y}^*; X_\ell \boldsymbol{\beta}_\ell, \delta\sigma_\ell^2 I_{n^*}) f_{N_{n^*}}(\mathbf{y}^*; X_\ell \overline{\overline{\boldsymbol{\beta}}}_0, \delta\sigma_0^2 I_{n^*})\, d\mathbf{y}^*, \tag{8}$$

where $\overline{\overline{\boldsymbol{\beta}}}_0 = (\boldsymbol{\beta}_0^T, \mathbf{0}_{d_\ell - d_0}^T)^T$ and $\mathbf{0}_k$ being a vector of zeros of length k. The product of the two normal densities involved in the integrand is given by

$$\begin{aligned} &f_{N_{n^*}}(\mathbf{y}^*; X_\ell \boldsymbol{\beta}_\ell, \delta\sigma_\ell^2 I_{n^*}) f_{N_{n^*}}(\mathbf{y}^*; X_\ell \overline{\overline{\boldsymbol{\beta}}}_0, \delta\sigma_0^2 I_{n^*}) \\ &\quad = (2\pi)^{-(n^*-d_\ell)/2} [\delta(\sigma_0^2 + \sigma_\ell^2)]^{-(n^*-d_\ell)/2} \\ &\quad\quad \times |X_\ell^{*T} X_\ell^*|^{-1/2} f_{N_{n^*}}(\mathbf{y}^*; E^{-1}D, E^{-1}) \\ &\quad\quad \times f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell; \overline{\overline{\boldsymbol{\beta}}}_0, \delta(\sigma_\ell^2 + \sigma_0^2)(X_\ell^{*T} X_\ell^*)^{-1}) \end{aligned} \tag{9}$$

with

$$\begin{aligned} &E = \left(\frac{\sigma_\ell^2 + \sigma_0^2}{\delta\sigma_0^2\sigma_\ell^2}\right) I_{n^*} \quad \text{and} \\ &D = \frac{1}{\delta\sigma_0^2} X_\ell^* \overline{\overline{\boldsymbol{\beta}}}_0 + \frac{1}{\delta\sigma_\ell^2} X_\ell^* \boldsymbol{\beta}_\ell = \frac{1}{\delta} X_\ell^* \left(\frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_0^2} \overline{\overline{\boldsymbol{\beta}}}_0 + \frac{\sigma_0^2}{\sigma_\ell^2 + \sigma_0^2} \boldsymbol{\beta}_\ell\right). \end{aligned} \tag{10}$$

Note that (9) was obtained using the property

$$\begin{aligned} &f_{N_n}(\mathbf{y}; M\boldsymbol{\xi}_1, A_1) f_{N_n}(\mathbf{y}; M\boldsymbol{\xi}_2, A_2) \\ &\quad = (2\pi)^{-(n-p)/2} |A_1 + A_2|^{-1/2} |M^T (A_1 + A_2)^{-1} M|^{-1/2} \\ &\quad\quad \times f_{N_n}(\mathbf{y}; E_1^{-1} D_1, E_1^{-1}) f_{N_n}(\boldsymbol{\xi}_1; \boldsymbol{\xi}_2, A_1 + A_2) \end{aligned} \tag{11}$$

with

$$E_1 = A_1^{-1} + A_2^{-1} \quad \text{and} \quad D_1 = A_1^{-1} M\boldsymbol{\xi}_1 + A_2^{-1} M\boldsymbol{\xi}_2.$$

In (11), M is a $n \times p$ matrix of rank p ($p \le n$), $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are vectors of length p and A_1 and A_2 are positive definite matrices of dimension $n \times n$. Expression (11) can be easily obtained using the identity:

$$\begin{aligned} &(\mathbf{y} - M\boldsymbol{\xi}_1)^T A_1^{-1} (\mathbf{y} - M\boldsymbol{\xi}_1) + (\mathbf{y} - M\boldsymbol{\xi}_2)^T A_2^{-1} (\mathbf{y} - M\boldsymbol{\xi}_2) \\ &\quad = \mathbf{y}^T E\mathbf{y} - 2\mathbf{y}^T (A_1^{-1} M\boldsymbol{\xi}_1 + A_2^{-1} M\boldsymbol{\xi}_2) + \boldsymbol{\xi}_1^T \\ &\quad\quad + M^T A_1^{-1} M\boldsymbol{\xi}_1 + \boldsymbol{\xi}_2^T + M^T A_2^{-1} M\boldsymbol{\xi}_2 \\ &\quad = [C^T \mathbf{y} - C^{-1} D]^T [C^T \mathbf{y} - C^{-1} D] \\ &\quad\quad + (\boldsymbol{\xi}_2 - \boldsymbol{\xi}_1)^T M^T (A_1 + A_2)^{-1} M(\boldsymbol{\xi}_2 - \boldsymbol{\xi}_1), \end{aligned}$$

with C being a $n \times n$ lower triangular matrix (the Cholesky decomposition) with nonzero elements in the diagonal such that $\mathrm{E}_1 = \mathrm{C}\mathrm{C}^T$.

Replacing (9) in (8), we obtain

$$\begin{aligned}
&\pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{\beta}_0, \sigma_0^2; \mathrm{X}_\ell^*, \delta) \\
&\quad = \frac{\pi^{-(d_\ell - n^*)/2}}{\sigma_\ell^2 \Gamma((n^* - d_\ell)/2)} \big|\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*\big|^{1/2} (2\pi)^{-(n^* - d_\ell)/2} \\
&\qquad \times \big[\delta(\sigma_0^2 + \sigma_\ell^2)\big]^{-(n^* - d_\ell)/2} \big|\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*\big|^{-1/2} \\
&\qquad \times f_{N_{n^*}}\big(\boldsymbol{\beta}_\ell; \overline{\overline{\boldsymbol{\beta}}}_0, \delta(\sigma_\ell^2 + \sigma_0^2)(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\big) \\
&\qquad \times \int \big(\mathbf{y}^{*T}\mathrm{P}_\ell^*\mathbf{y}^*\big)^{(n^* - d_\ell)/2} f_{N_{n^*}}\big(\mathbf{y}^*; \mathrm{E}^{-1}\mathrm{D}, \mathrm{E}^{-1}\big)\, d\mathbf{y}^*,
\end{aligned} \tag{12}$$

with E and D given in (10).

We set

$$\mathbf{z} = \mathrm{E}^{1/2}\big(\mathbf{y}^* - \mathrm{E}^{-1}\mathrm{D}\big) = \zeta^{1/2}\big(\mathbf{y}^* - \mathrm{X}_\ell^*\Gamma\big),$$

where $\zeta = (\frac{\sigma_\ell^2 + \sigma_0^2}{\delta\sigma_0^2\sigma_\ell^2})$ and $\Gamma = (\zeta\delta)^{-1}(\frac{\sigma_\ell^2}{\sigma_\ell^2 + \sigma_0^2}\overline{\overline{\boldsymbol{\beta}}}_0 + \frac{\sigma_0^2}{\sigma_\ell^2 + \sigma_0^2}\boldsymbol{\beta}_\ell)$. Therefore we have $\mathbf{y}^* = \zeta^{-1/2}\mathbf{z} + \mathrm{X}_\ell^*\Gamma$, $d\mathbf{y}^* = \zeta^{-n^*/2}\, d\mathbf{z}$ and

$$f_{N_{n^*}}\big(\mathbf{y}^*; \mathrm{E}^{-1}\mathrm{D}, \mathrm{E}^{-1}\big)\, d\mathbf{y}^* = f_{N_{n^*}}(\mathbf{z}; \mathbf{0}_{n^*}, \mathrm{I}_{n^*})\, d\mathbf{z}$$

since the term $\zeta^{-n^*/2}$, coming from the Jacobian of the transformation, cancels out with the determinant of the variance, that is $|\mathrm{E}|^{1/2} = \zeta^{n^*/2}$. Moreover,

$$\begin{aligned}
\mathbf{y}^{*T}\mathrm{P}_\ell^*\mathbf{y}^* &= \big(\zeta^{-1/2}\mathbf{z} + \mathrm{X}_\ell^*\Gamma\big)^T \mathrm{P}_\ell^*\big(\zeta^{-1/2}\mathbf{z} + \mathrm{X}_\ell^*\Gamma\big) \\
&= \zeta^{-1}\mathbf{z}^T\mathrm{P}_\ell^*\mathbf{z} + \zeta^{-1/2}\mathbf{z}^T\mathrm{P}_\ell^*\mathrm{X}_\ell^*\Gamma \\
&\quad + \Gamma^T\mathrm{X}_\ell^{*T}\mathrm{P}_\ell^*\zeta^{-1/2}\mathbf{z} + \Gamma^T\mathrm{X}_\ell^{*T}\mathrm{P}_\ell^*\mathrm{X}_\ell^*\Gamma \\
&= \zeta^{-1}\mathbf{z}^T\mathrm{P}_\ell^*\mathbf{z}
\end{aligned} \tag{13}$$

since $\mathrm{X}_\ell^{*T}\mathrm{P}_\ell^* = \mathrm{P}_\ell^*\mathrm{X}_\ell^* = \mathbf{0}$.

Returning back to (12), we obtain

$$\begin{aligned}
&\pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{\beta}_0, \sigma_0^2; \mathrm{X}_\ell^*, \delta) \\
&\quad = 2^{-(n^* - d_\ell)/2}\left[\sigma_\ell^2\Gamma\left(\frac{n^* - d_\ell}{2}\right)\right]^{-1}\big[\delta(\sigma_0^2 + \sigma_\ell^2)\big]^{-(n^* - d_\ell)/2} \\
&\qquad \times f_{N_{n^*}}\big(\boldsymbol{\beta}_\ell; \overline{\overline{\boldsymbol{\beta}}}_0, \delta(\sigma_\ell^2 + \sigma_0^2)(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\big) \\
&\qquad \times \zeta^{-(n^* - d_\ell)/2}\int \big(\mathbf{z}^T\mathrm{P}_\ell^*\mathbf{z}\big)^{(n^* - d_\ell)/2} f_{N_{n^*}}(\mathbf{z}; \mathbf{0}_{n^*}, \mathrm{I}_{n^*})\, d\mathbf{z}
\end{aligned}$$

$$
\begin{aligned}
&= 2^{-(n^*-d_\ell)/2}\left[\Gamma\left(\frac{n^*-d_\ell}{2}\right)\right]^{-1}\left[\delta\left(\sigma_0^2+\sigma_\ell^2\right)\right]^{-(n^*-d_\ell)/2}\delta^{(n^*-d_\ell)/2}\\
&\quad\times\left(\sigma_0^2\right)^{(n^*-d_\ell)/2}\left(\sigma_\ell^2\right)^{(n^*-d_\ell)/2-1}\left(\sigma_0^2+\sigma_\ell^2\right)^{-(n^*-d_\ell)/2}\\
&\quad\times f_{N_{n^*}}\left(\boldsymbol{\beta}_\ell;\overline{\overline{\boldsymbol{\beta}}}_0,\delta\left(\sigma_\ell^2+\sigma_0^2\right)\left(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*\right)^{-1}\right)\\
&\quad\times E\left[\left(\mathbf{z}^T\mathrm{P}_\ell^*\mathbf{z}\right)^{(n^*-d_\ell)/2}\right]\\
&= 2^{-(n^*-d_\ell)/2}\left[\Gamma\left(\frac{n^*-d_\ell}{2}\right)\right]^{-1}\left(\sigma_0^2\right)^{(n^*-d_\ell)/2}\\
&\quad\times\left(\sigma_\ell^2\right)^{(n^*-d_\ell)/2-1}\left(\sigma_0^2+\sigma_\ell^2\right)^{-(n^*-d_\ell)}\\
&\quad\times f_{N_{n^*}}\left(\boldsymbol{\beta}_\ell;\overline{\overline{\boldsymbol{\beta}}}_0,\delta\left(\sigma_\ell^2+\sigma_0^2\right)\left(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*\right)^{-1}\right)2^{(n^*-d_\ell)/2}\\
&\quad\times\frac{\Gamma((n^*-d_\ell)/2+(n^*-d_\ell)/2)}{\Gamma((n^*-d_\ell)/2)},
\end{aligned}
$$

since

$$E\left[\left(\mathbf{x}^T K\mathbf{x}\right)^h\right]=2^h\frac{\Gamma(h+r/2)}{r/2},$$

where $h>0$, K is a $n\times n$ symmetric and idempotent matrix of rank r, $\mathbf{x}\sim N_n(\mathbf{0}_n,\mathrm{I}_n)$ and, therefore, $\mathbf{x}^T K\mathbf{x}\sim\chi_r^2$.

Thus, (14) becomes

$$
\begin{aligned}
&\pi_\ell^{\text{J-PEP}}\left(\boldsymbol{\beta}_\ell,\sigma_\ell^2|\boldsymbol{\beta}_0,\sigma_0^2;\mathrm{X}_\ell^*,\delta\right)\\
&\quad=\frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2}\left(\sigma_0^2\right)^{-(n^*-d_\ell)/2}\left(\sigma_\ell^2\right)^{(n^*-d_\ell)/2-1}\left(1+\frac{\sigma_\ell^2}{\sigma_0^2}\right)^{-(n^*-d_\ell)}\\
&\qquad\times f_{N_{d_\ell}}\left(\boldsymbol{\beta}_\ell;\overline{\overline{\boldsymbol{\beta}}}_0,\delta\left(\sigma_\ell^2+\sigma_0^2\right)\left(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*\right)^{-1}\right).
\end{aligned}
\tag{14}
$$

3 The J-PEP Bayes factor

The Bayes factor of any model M_ℓ ($\ell\neq 0$) versus the reference model M_0, under the J-PEP prior approach, is given by

$$BF_{\ell 0}^{\text{J-PEP}}=\frac{\int f_{N_n}(\mathbf{y};\mathrm{X}_\ell\boldsymbol{\beta}_\ell,\sigma_\ell^2\mathrm{I}_n)\pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell,\sigma_\ell^2|\mathrm{X}_\ell^*,\delta)\,d\boldsymbol{\beta}_\ell\,d\sigma_\ell^2}{\int f_{N_n}(\mathbf{y};\mathrm{X}_0\boldsymbol{\beta}_0,\sigma_0^2\mathrm{I}_n)\pi_0^N(\boldsymbol{\beta}_0,\sigma_0^2|\mathrm{X}_0^*)\,d\boldsymbol{\beta}_0\,d\sigma_0^2}$$

with the denominator given by

$$m_0^N(\mathbf{y}|\mathrm{X}_0)=c_0\pi^{(d_0-n)/2}\left|\mathrm{X}_0^T\mathrm{X}_0\right|^{-1/2}\Gamma\left(\frac{n-d_0}{2}\right)RSS_0^{-(n-d_0)/2}.$$

Using (14), the numerator is given by

$$
\begin{aligned}
& m_\ell^{\text{J-PEP}}(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \\
&\quad = \iiiint f_{N_n}(\mathbf{y}; \mathrm{X}_\ell \boldsymbol{\beta}_\ell, \sigma_\ell^2 \mathrm{I}_n) \pi_\ell^{\text{J-PEP}}(\boldsymbol{\beta}_\ell, \sigma_\ell^2 | \boldsymbol{\beta}_0, \sigma_0^2; \mathrm{X}_\ell^*, \delta) \\
&\qquad \times \pi_0^N(\boldsymbol{\beta}_0, \sigma_0^2 | \mathrm{X}_0^*) \, d\boldsymbol{\beta}_\ell \, d\sigma_\ell^2 \, d\boldsymbol{\beta}_0 \, d\sigma_0^2 \\
&\quad = \iiiint \frac{c_0}{\sigma_0^2} C_\ell f_{N_n}(\mathbf{y}; \mathrm{X}_\ell \boldsymbol{\beta}_\ell, \sigma_\ell^2 \mathrm{I}_n) \\
&\qquad \times f_{N_{d_\ell}}(\boldsymbol{\beta}_\ell; \overline{\overline{\boldsymbol{\beta}}}_0, \delta(\sigma_\ell^2 + \sigma_0^2)(\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1}) \, d\boldsymbol{\beta}_\ell \, d\sigma_\ell^2 \, d\boldsymbol{\beta}_0 \, d\sigma_0^2,
\end{aligned}
$$

with

$$
C_\ell = (\sigma_0^2)^{-(n^*-d_\ell)/2} (\sigma_\ell^2)^{(n^*-d_\ell)/2-1} \left(1 + \frac{\sigma_\ell^2}{\sigma_0^2}\right)^{-(n^*-d_\ell)} \frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2}. \tag{15}
$$

Integrating out $\boldsymbol{\beta}_\ell$, we obtain

$$
m_\ell^{\text{J-PEP}}(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = \iiint \frac{c_0}{\sigma_0^2} C_\ell [f_{N_n}(\mathbf{y}; \mathrm{X}_\ell \overline{\overline{\boldsymbol{\beta}}}_0, \Sigma'_\ell)] \, d\boldsymbol{\beta}_0 \, d\sigma_\ell^2 \, d\sigma_0^2,
$$

with

$$
\Sigma'_\ell = \sigma_\ell^2 \mathrm{I}_n + \delta(\sigma_\ell^2 + \sigma_0^2) \mathrm{X}_\ell (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^T.
$$

The above expression was obtained using the following formula:

$$
\int f_{N_n}(\mathbf{y}; \mathrm{M}\boldsymbol{\xi}_1, \mathrm{A}_1) f_{N_p}(\boldsymbol{\xi}_1; \boldsymbol{\xi}_2, \mathrm{A}_3) \, d\boldsymbol{\xi}_1 = f_{N_n}(\mathbf{y}; \mathrm{M}\boldsymbol{\xi}_2, \mathrm{A}_1 + \mathrm{M}\mathrm{A}_3\mathrm{M}^T),
$$

with M being a $n \times p$ matrix of rank p ($p \le n$), $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ being vectors of length p and A_1 and A_3 being positive definite matrices of dimensions $n \times n$ and $p \times p$, respectively.

Moreover,

$$
\begin{aligned}
m_\ell^{\text{J-PEP}}(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) &= \iiint \frac{c_0}{\sigma_0^2} C_\ell [f_{N_n}(\mathbf{y}; \mathrm{X}_\ell \overline{\overline{\boldsymbol{\beta}}}_0, \Sigma'_\ell)] \, d\boldsymbol{\beta}_0 \, d\sigma_\ell^2 \, d\sigma_0^2 \\
&= \iiint \frac{c_0}{\sigma_0^2} C_\ell [f_{N_n}(\mathbf{y}; \mathrm{X}_0 \boldsymbol{\beta}_0, \Sigma'_\ell)] \, d\boldsymbol{\beta}_0 \, d\sigma_\ell^2 \, d\sigma_0^2 \\
&= \iint \frac{c_0}{\sigma_0^2} C_\ell \Big[(2\pi)^{-(n-d_0)/2} |\Sigma'_\ell|^{-1/2} |\mathrm{X}_0^T \Sigma_\ell'^{-1} \mathrm{X}_0|^{-1/2} \\
&\qquad \times \exp\left\{ -\frac{1}{2} \mathbf{y}^T A_\Sigma \mathbf{y} \right\} \Big] \, d\sigma_\ell^2 \, d\sigma_0^2,
\end{aligned}
$$

where

$$
A_\Sigma = \Sigma_\ell'^{-1} - \Sigma_\ell'^{-1} \mathrm{X}_0 [\mathrm{X}_0^T \Sigma_\ell'^{-1} \mathrm{X}_0]^{-1} \mathrm{X}_0^T \Sigma_\ell'^{-1},
$$

since

$$\int f_{N_n}(\mathbf{y}; \mathrm{M}\boldsymbol{\xi}_1, \mathrm{A}_1)\, d\boldsymbol{\xi}_1 = (2\pi)^{-(n-p)/2} |\mathrm{A}_1|^{-1/2} |\mathrm{M}^T \mathrm{A}_1^{-1} \mathrm{M}|^{-1/2} \times \exp\left\{-\frac{1}{2}\mathbf{y}^T [\mathrm{A}_1^{-1} - \mathrm{A}_1^{-1}\mathrm{M}(\mathrm{M}^T \mathrm{A}_1^{-1}\mathrm{M})^{-1}\mathrm{M}^T \mathrm{A}_1^{-1}]\mathbf{y}\right\}$$

with M being a $n \times p$ matrix of rank p ($p \leq n$), $\boldsymbol{\xi}_1$ being a vector of length p and A_1 being a positive definite matrix of dimension $n \times n$.

Substituting expression (15), we obtain

$$\begin{aligned}
m_\ell^{\text{J-PEP}}&(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \\
&= \iint \frac{c_0}{\sigma_0^2} (\sigma_0^2)^{-(n^*-d_\ell)/2} (\sigma_\ell^2)^{(n^*-d_\ell)/2-1} \left(1 + \frac{\sigma_\ell^2}{\sigma_0^2}\right)^{-(n^*-d_\ell)} \frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2} \\
&\quad \times \left[(2\pi)^{-(n-d_0)/2} |\Sigma_\ell'|^{-1/2} |\mathrm{X}_0^T \Sigma_\ell'^{-1} \mathrm{X}_0|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}^T A_\Sigma \mathbf{y}\right\}\right] d\sigma_\ell^2\, d\sigma_0^2 \\
&= c_0 (2\pi)^{-(n-d_0)/2} \frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2} \\
&\quad \times \iint (\sigma_0^2)^{-2} \left(\frac{\sigma_\ell^2}{\sigma_0^2}\right)^{(n^*-d_\ell)/2-1} \left(1 + \frac{\sigma_\ell^2}{\sigma_0^2}\right)^{-(n^*-d_\ell)} \\
&\quad \times |\Sigma_\ell'|^{-1/2} |\mathrm{X}_0^T \Sigma_\ell'^{-1} \mathrm{X}_0|^{-1/2} \exp\left\{-\frac{1}{2}\mathbf{y}^T A_\Sigma \mathbf{y}\right\} d\sigma_\ell^2\, d\sigma_0^2.
\end{aligned} \tag{16}$$

We now set

$$r = \sqrt{\sigma_0^2 + \sigma_\ell^2} \quad \text{and} \quad \phi = \arctan\left(\sqrt{\frac{\sigma_\ell^2}{\sigma_0^2}}\right)$$

for $r \in [0, +\infty)$ and $\phi \in [0, \pi/2]$. The inverse transformations are given by

$$\sigma_0^2 = r^2 \cos^2\phi \quad \text{and} \quad \sigma_\ell^2 = r^2 \sin^2\phi \tag{17}$$

while the Jacobian is

$$\begin{aligned}
J(r,\phi) &= \begin{vmatrix} \frac{\partial \sigma_0^2}{\partial r} & \frac{\partial \sigma_0^2}{\partial \phi} \\ \frac{\partial \sigma_\ell^2}{\partial r} & \frac{\partial \sigma_\ell^2}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \frac{\partial (r^2 \cos^2\phi)}{\partial r} & \frac{(\partial r^2 \cos^2\phi)}{\partial \phi} \\ \frac{\partial (r^2 \sin^2\phi)}{\partial r} & \frac{(\partial r^2 \sin^2\phi)}{\partial \phi} \end{vmatrix} \\
&= \begin{vmatrix} 2r\cos^2\phi & -2r^2 \cos\phi \sin\phi \\ 2r \sin^2\phi & 2r^2 \sin\phi\cos\phi \end{vmatrix} \\
&= 4r^3 \sin\phi\cos\phi(\cos^2\phi + \sin^2\phi) = 4r^3 \sin\phi\cos\phi.
\end{aligned} \tag{18}$$

Then the matrix Σ'_ℓ becomes equal to

$$\begin{aligned}\Sigma'_\ell &= \sigma_\ell^2 \mathrm{I}_n + \delta(\sigma_\ell^2 + \sigma_0^2)\mathrm{X}_\ell(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\mathrm{X}_\ell^T \\ &= r^2 \sin^2\phi \mathrm{I}_n + r^2\delta\mathrm{X}_\ell(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\mathrm{X}_\ell^T \\ &= r^2 B(\phi) \end{aligned} \tag{19}$$

with $B(\phi)$ being a $n \times n$ matrix given by

$$B(\phi) = \sin^2\phi \mathrm{I}_n + \delta\mathrm{X}_\ell(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\mathrm{X}_\ell^T \tag{20}$$

while A_Σ can be rewritten as

$$\begin{aligned}A_\Sigma &= \Sigma_\ell'^{-1} - \Sigma_\ell'^{-1}\mathrm{X}_0\big[\mathrm{X}_0^T\Sigma_\ell'^{-1}\mathrm{X}_0\big]^{-1}\mathrm{X}_0^T\Sigma_\ell'^{-1} \\ &= r^{-2}B^{-1}(\phi) - r^{-2}B^{-1}(\phi)\mathrm{X}_0\big[\mathrm{X}_0^T r^{-2}B^{-1}(\phi)\mathrm{X}_0\big]^{-1}\mathrm{X}_0^T r^{-2}B^{-1}(\phi) \\ &= r^{-2}\big[B^{-1}(\phi) - B^{-1}(\phi)\mathrm{X}_0 A^{-1}(\phi)\mathrm{X}_0^T B^{-1}(\phi)\big]\end{aligned}$$

with

$$A(\phi) = \mathrm{X}_0^T B^{-1}(\phi)\mathrm{X}_0 \tag{21}$$

being a $d_0 \times d_0$ matrix. Moreover, we have that

$$\mathbf{y}^T A_\Sigma \mathbf{y} = r^{-2}D(\phi) \tag{22}$$

with

$$D(\phi) = \mathbf{y}^T\big[B^{-1}(\phi) - B^{-1}(\phi)\mathrm{X}_0 A^{-1}(\phi)\mathrm{X}_0^T B^{-1}(\phi)\big]\mathbf{y} \tag{23}$$

being a scalar. Finally, the first three terms in the integrand of (16) can be written as

$$\begin{aligned}&(\sigma_0^2)^{-2}\left(\frac{\sigma_\ell^2}{\sigma_0^2}\right)^{(n^*-d_\ell)/2-1}\left(1+\frac{\sigma_\ell^2}{\sigma_0^2}\right)^{-(n^*-d_\ell)} \\ &\quad= \big(r^2\cos^2\phi\big)^{-2}\left(\frac{\sin^2\phi}{\cos^2\phi}\right)^{(n^*-d_\ell)/2-1} \\ &\qquad\times\left(\frac{r^2\cos^2\phi + r^2\sin^2\phi}{r^2\cos^2\phi}\right)^{-(n^*-d_\ell)} \\ &\quad= \big(r^2\cos^2\phi\big)^{-2}\left(\frac{\sin^2\phi}{\cos^2\phi}\right)^{(n^*-d_\ell)/2-1}\big(\cos^2\phi\big)^{n^*-d_\ell} \\ &\quad= r^{-4}(\sin\phi\cos\phi)^{n^*-d_\ell-2}.\end{aligned} \tag{24}$$

Using the transformation (17) and the corresponding Jacobian given by (18), as well as expressions (19), (22) and (24), the marginal likelihood (16) now becomes

$$
\begin{aligned}
&m_\ell^{\text{J-PEP}}(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \\
&\quad = c_0(2\pi)^{-(n-d_0)/2} \frac{\Gamma(n^* - d_\ell)}{\Gamma((n^* - d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \int_0^{\infty} \frac{r^{-4}(\sin\phi\cos\phi)^{n^*-d_\ell-2}}{|r^2 B(\phi)|^{1/2} |r^{-2}\mathrm{X}_0^T B^{-1}(\phi)\mathrm{X}_0|^{1/2}} \\
&\qquad \times \exp\left\{-\frac{1}{2} r^{-2} D(\phi)\right\} 4r^3 \sin\phi\cos\phi \, dr\, d\phi \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2} \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2} |\mathrm{X}_0^T B^{-1}(\phi)\mathrm{X}_0|^{1/2}} \\
&\qquad \times \int_0^{\infty} r^{-n+d_0-1} \exp\left\{-\frac{1}{2} r^{-2} D(\phi)\right\} dr\, d\phi.
\end{aligned} \tag{25}
$$

We now set $w = 1/r$ ($\Leftrightarrow r = w^{-1}$ and $dr = (-1)w^{-2}\,dw$), resulting in

$$
\begin{aligned}
&m_\ell^{\text{J-PEP}}(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2} \frac{\Gamma(n^* - d_\ell)}{\Gamma((n^* - d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}} \int_0^{\infty} w^{n-d_0+1} \exp\left\{-\frac{1}{2} w^2 D(\phi)\right\} w^{-2}\, dw\, d\phi \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2} \frac{\Gamma(n^* - d_\ell)}{\Gamma((n^* - d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2} D(\phi)} \\
&\qquad \times \int_0^{\infty} w^{n-d_0-2} \frac{w}{D(\phi)^{-1}} \exp\left\{-\frac{w^2}{2D(\phi)^{-1}}\right\} dw\, d\phi \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2} \frac{\Gamma(n^* - d_\ell)}{\Gamma((n^* - d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2} D(\phi)} \int_0^{\infty} w^{n-d_0-2} f_R\big(w; D(\phi)^{-1}\big)\, dw\, d\phi \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2} \frac{\Gamma(n^* - d_\ell)}{\Gamma((n^* - d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2} D(\phi)} E_R\big(w^{n-d_0-2}; D(\phi)^{-1}\big)\, d\phi,
\end{aligned}
$$

where $f_R(w; s^2)$ is the density function of the Rayleigh distribution with scale parameter s^2 (which here is equal to $D(\phi)^{-1}$) and variance $s^2(4-\pi)/2$. Moreover, by $E_R(w^k; s^2)$ we denote the corresponding kth moment about zero which is given by $s^k 2^{k/2}\Gamma(1+k/2)$. Therefore, we have:

$$
\begin{aligned}
&m_\ell^{\text{J-PEP}}(\mathbf{y}|\mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2}\frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}2^{(n-d_0-2)/2}\Gamma(1+(n-d_0-2)/2)}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}[D(\phi)]^{1+(n-d_0-2)/2}}\,d\phi \\
&\quad = 4c_0(2\pi)^{-(n-d_0)/2}\frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2}2^{(n-d_0)/2-1}\Gamma\left(\frac{n-d_0}{2}\right) \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}[D(\phi)]^{(n-d_0)/2}}\,d\phi \\
&\quad = 2c_0\pi^{-(n-d_0)/2}\frac{\Gamma(n^*-d_\ell)\Gamma((n-d_0)/2)}{\Gamma((n^*-d_\ell)/2)^2} \\
&\qquad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}[D(\phi)]^{(n-d_0)/2}}\,d\phi.
\end{aligned}
$$

Hence, the Bayes factor of model M_ℓ ($\ell \neq 0$) versus the reference model M_0, under the J-PEP prior approach, is given by

$$
\begin{aligned}
BF_{\ell 0}^{\text{J-PEP}} &= \frac{2c_0\pi^{-(n-d_0)/2}((\Gamma(n^*-d_\ell)\Gamma((n-d_0)/2))/(\Gamma((n^*-d_\ell)/2)^2))}{c_0\pi^{(d_0-n)/2}|\mathrm{X}_0^T\mathrm{X}_0|^{-1/2}\Gamma((n-d_0)/2)RSS_0^{-(n-d_0)/2}} \\
&\quad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}[D(\phi)]^{(n-d_0)/2}}\,d\phi \\
&= 2\frac{\Gamma(n^*-d_\ell)}{\Gamma((n^*-d_\ell)/2)^2}\left|\mathrm{X}_0^T\mathrm{X}_0\right|^{1/2}RSS_0^{(n-d_0)/2} \\
&\quad \times \int_0^{\pi/2} \frac{(\sin\phi\cos\phi)^{n^*-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}[D(\phi)]^{(n-d_0)/2}}\,d\phi.
\end{aligned} \tag{26}
$$

Under the J-PEP approach, we set $(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*) = (\mathrm{X}_\ell^T\mathrm{X}_\ell)$, $n^* = n$ and $\delta = n$, and thus

$$
\begin{aligned}
B(\phi) &= \sin^2\phi\mathrm{I}_n + \delta\mathrm{X}_\ell\left(\mathrm{X}_\ell^T\mathrm{X}_\ell\right)^{-1}\mathrm{X}_\ell^T \\
&= \sin^2\phi\mathrm{I}_n + \delta\mathrm{H}_\ell.
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
B^{-1}(\phi) &= \left[\sin^2\phi \mathrm{I}_n + \delta \mathrm{H}_\ell\right]^{-1} = \frac{1}{\sin^2\phi}\left[\mathrm{I}_n + \frac{\delta}{\sin^2\phi}\mathrm{X}_\ell(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}\mathrm{X}_\ell^T\right]^{-1} \\
&= \frac{1}{\sin^2\phi}\left[\mathrm{I}_n^{-1} - \mathrm{I}_n^{-1}\frac{\delta}{\sin^2\phi}\mathrm{X}_\ell\left([(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}]^{-1} + \frac{\delta}{\sin^2\phi}\mathrm{X}_\ell^T\mathrm{X}_\ell\right)^{-1}\mathrm{X}_\ell^T\mathrm{I}_n^{-1}\right] \\
&= \frac{1}{\sin^2\phi}\left[\mathrm{I}_n - \frac{\delta}{\sin^2\phi}\frac{\sin^2\phi}{\delta + \sin^2\phi}\mathrm{H}_\ell\right] \\
&= \frac{1}{\sin^2\phi}\left[\mathrm{I}_n - \frac{\delta}{\delta + \sin^2\phi}\mathrm{H}_\ell\right] \\
&= \frac{1}{\sin^2\phi}\frac{\delta}{\delta + \sin^2\phi}[\mathrm{I}_n - \mathrm{H}_\ell] + \frac{1}{\sin^2\phi}\frac{\sin^2\phi}{\delta + \sin^2\phi}\mathrm{I}_n \\
&= \frac{\delta}{\sin^2\phi(\delta + \sin^2\phi)}\mathrm{P}_\ell + \frac{1}{\delta + \sin^2\phi}\mathrm{I}_n \qquad (27)
\end{aligned}
$$

and $|B(\phi)| = |\sin^2\phi \mathrm{I}_n + \delta\mathrm{H}_\ell| = (\sin^2\phi)^n|\mathrm{I}_n + \frac{\delta}{\sin^2\phi}\mathrm{H}_\ell| = (\sin^2\phi)^n|\mathrm{I}_{d_\ell} + \frac{\delta}{\sin^2\phi} \times (\mathrm{X}_\ell^T\mathrm{X}_\ell)(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}|$ resulting in

$$
|B(\phi)| = (\sin^2\phi)^n\left(1 + \frac{\delta}{\sin^2\phi}\right)^{d_\ell} = (\sin^2\phi)^{n-d_\ell}(\delta + \sin^2\phi)^{d_\ell}.
$$

Also $\mathbf{y}^T B^{-1}(\phi)\mathbf{y} = \frac{\delta}{\sin^2\phi(\delta+\sin^2\phi)}\mathbf{y}^T[\mathrm{I}_n - \mathrm{H}_\ell]\mathbf{y} + \frac{1}{\delta+\sin^2\phi}\mathbf{y}^T\mathbf{y} = \frac{1}{\delta+\sin^2\phi} \times (\frac{\delta}{\sin^2\phi}RSS_\ell + \mathbf{y}^T\mathbf{y})$. From (21), $A(\phi)$ is now given by

$$
\begin{aligned}
A(\phi) = \mathrm{X}_0^T B^{-1}(\phi)\mathrm{X}_0 &= \frac{1}{\sin^2\phi}\mathrm{X}_0^T\left[\mathrm{I}_n - \frac{\delta}{\delta + \sin^2\phi}\mathrm{H}_\ell\right]\mathrm{X}_0 \\
&= \frac{1}{\sin^2\phi}\left[\mathrm{X}_0^T\mathrm{X}_0 - \frac{\delta}{\delta + \sin^2\phi}\mathrm{X}_0^T\mathrm{H}_\ell\mathrm{X}_0\right] \\
&= \frac{1}{\sin^2\phi}\left[\mathrm{X}_0^T\mathrm{X}_0 - \frac{\delta}{\delta + \sin^2\phi}\mathrm{X}_0^T\mathrm{X}_0\right] \\
&= \frac{1}{\delta + \sin^2\phi}\mathrm{X}_0^T\mathrm{X}_0
\end{aligned}
$$

since H_ℓ is idempotent and $\mathrm{X}_0^T\mathrm{H}_\ell = \mathrm{X}_0$ for any model M_0 nested in M_ℓ. This comes from the blockwise formula where for any $X_\ell = [X_0, X_{\ell\setminus 0}]$ we have

$$
\begin{aligned}
H_\ell &= H_0 + H_{(I_n - H_0)X_{\ell\setminus 0}} \quad \Leftrightarrow \\
X_0^T H_\ell &= X_0^T H_0 + X_0^T H_{P_0 X_{\ell\setminus 0}}
\end{aligned}
$$

$$
\begin{aligned}
&= X_0^T + X_0^T \mathrm{P}_0 X_{\ell\setminus 0}\{[\mathrm{P}_0 X_{\ell\setminus 0}]^T \mathrm{P}_0 X_{\ell\setminus 0}\}^{-1}[\mathrm{P}_0 X_{\ell\setminus 0}]^T \\
&= X_0^T + (X_0^T - X_0^T H_0) X_{\ell\setminus 0}\{[\mathrm{P}_0 X_{\ell\setminus 0}]^T \mathrm{P}_0 X_{\ell\setminus 0}\}^{-1}[\mathrm{P}_0 X_{\ell\setminus 0}]^T = X_0^T.
\end{aligned}
$$

Therefore, $|A(\phi)| = (\delta + \sin^2\phi)^{-d_0}|\mathrm{X}_0^T\mathrm{X}_0|$ and $\mathrm{X}_0 A^{-1}(\phi)\mathrm{X}_0 = (\delta + \sin^2\phi) H_0$. From (23), we obtain that

$$
\begin{aligned}
D(\phi) =\ & \mathbf{y}^T B^{-1}(\phi)\mathbf{y} - \mathbf{y}^T B^{-1}(\phi)\mathrm{X}_0 A^{-1}(\phi)\mathrm{X}_0^T B^{-1}(\phi)\mathbf{y} \\
=\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + \mathbf{y}^T\mathbf{y}\right) \\
& - \mathbf{y}^T B^{-1}(\phi)\left[(\delta+\sin^2\phi)\mathrm{H}_0\right] B^{-1}(\phi)\mathbf{y} \\
=\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + \mathbf{y}^T\mathbf{y}\right) - (\delta+\sin^2\phi)\mathbf{y}^T \\
& \times \left[\frac{1}{\sin^2\phi}\left(\mathrm{I}_n - \frac{\delta}{\delta+\sin^2\phi}\mathrm{H}_\ell\right)\right] \\
& \times \mathrm{H}_0\left[\frac{1}{\sin^2\phi}\left(\mathrm{I}_n - \frac{\delta}{\delta+\sin^2\phi}\mathrm{H}_\ell\right)\right]\mathbf{y} \\
=\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + \mathbf{y}^T\mathbf{y}\right) \\
& - \frac{\delta+\sin^2\phi}{\sin^4\phi}\mathbf{y}^T\left(\mathrm{I}_n - \frac{\delta}{\delta+\sin^2\phi}\mathrm{H}_\ell\right)\mathrm{H}_0\left(\mathrm{I}_n - \frac{\delta}{\delta+\sin^2\phi}\mathrm{H}_\ell\right)\mathbf{y} \\
=\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + \mathbf{y}^T\mathbf{y}\right) \\
& - \frac{\delta+\sin^2\phi}{\sin^4\phi}\mathbf{y}^T\left(\mathrm{H}_0 - \frac{\delta}{\delta+\sin^2\phi}\mathrm{H}_\ell\mathrm{H}_0 - \frac{\delta}{\delta+\sin^2\phi}\mathrm{H}_0\mathrm{H}_\ell\right. \\
& \qquad\qquad \left. + \left[\frac{\delta}{\delta+\sin^2\phi}\right]^2 \mathrm{H}_\ell\mathrm{H}_0\mathrm{H}_\ell\right)\mathbf{y} \\
\overset{(\mathrm{H}_0\mathrm{H}_\ell=\mathrm{H}_0)}{=}\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + \mathbf{y}^T\mathbf{y}\right) - \frac{\delta+\sin^2\phi}{\sin^4\phi}\left[\frac{\sin^2\phi}{\delta+\sin^2\phi}\right]^2 \mathbf{y}^T\mathrm{H}_0\mathbf{y} \\
=\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathrm{H}_0\mathbf{y}\right) \\
=\ & \frac{1}{\delta+\sin^2\phi}\left(\frac{\delta}{\sin^2\phi} RSS_\ell + RSS_0\right).
\end{aligned}
$$

By substituting the above equations in (26), we obtain

$$
\begin{aligned}
BF_{\ell 0}^{\text{J-PEP}} &= 2\frac{\Gamma(n-d_\ell)}{\Gamma((n-d_\ell)/2)^2}\left|\mathrm{X}_0^T\mathrm{X}_0\right|^{1/2} RSS_0^{(n-d_0)/2} \\
&\quad\times\int_0^{\pi/2}\frac{(\sin\phi\cos\phi)^{n-d_\ell-1}}{|B(\phi)|^{1/2}|A(\phi)|^{1/2}[D(\phi)]^{(n-d_0)/2}}\,d\phi \\
&= 2\frac{\Gamma(n-d_\ell)}{\Gamma((n-d_\ell)/2)^2}\left|\mathrm{X}_0^T\mathrm{X}_0\right|^{1/2} RSS_0^{(n-d_0)/2} \\
&\quad\times\int_0^{\pi/2}\bigg((\sin\phi\cos\phi)^{n-d_\ell-1}\big(n+\sin^2\phi\big)^{(n-d_0)/2} \\
&\qquad\qquad\times\left(\frac{n}{\sin^2\phi}RSS_\ell + RSS_0\right)^{-(n-d_0)/2}\bigg) \\
&\quad\Big/\Big(\big(\sin^2\phi\big)^{(n-d_\ell)/2}\big(n+\sin^2\phi\big)^{d_\ell/2}\big(n+\sin^2\phi\big)^{-d_0/2}\left|\mathrm{X}_0^T\mathrm{X}_0\right|^{1/2}\Big)\,d\phi \\
&= 2\frac{\Gamma(n-d_\ell)}{\Gamma((n-d_\ell)/2)^2} \\
&\quad\times\int_0^{\pi/2}\bigg((\sin\phi\cos\phi)^{n-d_\ell-1}\big(n+\sin^2\phi\big)^{(n-d_0)/2} \\
&\qquad\qquad\times\big(\sin^2\phi\big)^{(n-d_0)/2}\left(n\frac{RSS_\ell}{RSS_0}+\sin^2\phi\right)^{-(n-d_0)/2}\bigg) \\
&\quad\Big/\Big(\big(\sin^2\phi\big)^{(n-d_\ell)/2}\big(n+\sin^2\phi\big)^{d_\ell/2}\big(n+\sin^2\phi\big)^{-d_0/2}\Big)\,d\phi \\
&= 2\frac{\Gamma(n-d_\ell)}{\Gamma((n-d_\ell)/2)^2} \\
&\quad\times\int_0^{\pi/2}\frac{(\sin\phi)^{n-d_0-1}(\cos\phi)^{n-d_\ell-1}(n+\sin^2\phi)^{(n-d_\ell)/2}}{(n(RSS_\ell/RSS_0)+\sin^2\phi)^{(n-d_0)/2}}\,d\phi.
\end{aligned}
\tag{28}
$$

For large n, we can write

$$
\begin{aligned}
\big(n+\sin^2\phi\big)^{(n-d_\ell)/2} &= \big(n+\sin^2\phi\big)^{n/2}\big(n+\sin^2\phi\big)^{-d_\ell/2} \\
&= n^{n/2}\left(1+\frac{\sin^2\phi/2}{n/2}\right)^{n/2}\big(n+\sin^2\phi\big)^{-d_\ell/2} \\
&\approx n^{n/2}\big(n+\sin^2\phi\big)^{-d_\ell/2}\exp\left(\frac{\sin^2\phi}{2}\right) \\
&\approx n^{(n-d_\ell)/2}\exp\left(\frac{\sin^2\phi}{2}\right).
\end{aligned}
$$

Similarly,

$$\begin{aligned}
&\left(n\frac{RSS_\ell}{RSS_0}+\sin^2\phi\right)^{(n-d_0)/2}\\
&\quad=\left[n\frac{RSS_\ell}{RSS_0}\right]^{(n-d_0)/2}\left(1+\frac{1/2\sin^2\phi\,(RSS_0/RSS_\ell)}{n/2}\right)^{n/2}\\
&\qquad\times\left(1+\frac{\sin^2\phi\,(RSS_0/RSS_\ell)}{n}\right)^{-d_0/2}\\
&\quad\approx\left[n\frac{RSS_\ell}{RSS_0}\right]^{(n-d_0)/2}\exp\left(\frac{1}{2}\sin^2\phi\frac{RSS_0}{RSS_\ell}\right).
\end{aligned}$$

Moreover, for large z we have

$$\log\Gamma(z)\approx\left(z-\frac{1}{2}\right)\log z-z+\frac{1}{2}\log(2\pi).$$

Hence,

$$\begin{aligned}
&\log\Gamma(n-d_\ell)\\
&\quad\approx\left(n-d_\ell-\frac{1}{2}\right)\log(n-d_\ell)-(n-d_\ell)+\frac{1}{2}\log(2\pi),
\end{aligned}$$

$$\begin{aligned}
&\log\Gamma\left(\frac{n-d_\ell}{2}\right)\\
&\quad\approx\left(\frac{n-d_\ell-1}{2}\right)\log\left(\frac{n-d_\ell}{2}\right)-\left(\frac{n-d_\ell}{2}\right)+\frac{1}{2}\log(2\pi),
\end{aligned}$$

$$\begin{aligned}
&\log\Gamma(n-d_\ell)-2\log\Gamma\left(\frac{n-d_\ell}{2}\right)\\
&\quad\approx\left(n-d_\ell-\frac{1}{2}\right)\log(n-d_\ell)-(n-d_\ell)+\frac{1}{2}\log(2\pi)\\
&\qquad-2\left(\frac{n-d_\ell-1}{2}\right)\log\left(\frac{n-d_\ell}{2}\right)\\
&\qquad+2\left(\frac{n-d_\ell}{2}\right)-2\frac{1}{2}\log(2\pi)\\
&\quad\approx\frac{1}{2}\log(n-d_\ell)-\frac{1}{2}\log(2\pi)+(n-d_\ell-1)\log 2\\
&\quad\approx\frac{1}{2}\log(n)+n\log 2.
\end{aligned}$$

From the above, we obtain that

$$\begin{aligned}
\log BF_{\ell 0}^{\text{J-PEP}}
&\approx \frac{1}{2}\log(n-d_\ell) - \frac{1}{2}\log(2\pi) + (n-d_\ell)\log 2 \\
&\quad + \log \int_0^{\pi/2} \frac{(\sin\phi)^{n-d_0-1}(\cos\phi)^{n-d_\ell-1} n^{(n-d_\ell)/2} \exp((\sin^2\phi)/2)}{[n(RSS_\ell/RSS_0)]^{(n-d_0)/2}\exp((1/2)\sin^2\phi(RSS_0/RSS_\ell))}\, d\phi \\
&\approx \frac{1}{2}\log(n-d_\ell) - \frac{1}{2}\log(2\pi) + (n-d_\ell)\log 2 \\
&\quad + \frac{n-d_\ell}{2}\log n - \frac{n-d_0}{2}\log n \log 2 - \frac{n-d_0}{2}\log\frac{RSS_\ell}{RSS_0} \\
&\quad + \log \int_0^{\pi/2} \frac{(\sin\phi)^{n-d_0-1}(\cos\phi)^{n-d_\ell-1}\exp((\sin^2\phi)/2)}{\exp((1/2)\sin^2\phi(RSS_0/RSS_\ell))}\, d\phi \\
&\approx \frac{1}{2}\log(n-d_\ell) - \frac{1}{2}\log(2\pi) + (n-d_\ell)\log 2 - \frac{d_\ell-d_0}{2}\log n \\
&\quad - \frac{n-d_0}{2}\log\frac{RSS_\ell}{RSS_0} \\
&\quad + \log \int_0^{\pi/2} \frac{(\sin\phi)^{n-d_0-1}(\cos\phi)^{n-d_\ell-1}\exp((\sin^2\phi)/2)}{\exp((1/2)\sin^2\phi(RSS_0/RSS_\ell))}\, d\phi \\
&\approx \frac{1}{2}\log n + n\log 2 - \frac{d_\ell-d_0}{2}\log n - \frac{n}{2}\log\frac{RSS_\ell}{RSS_0} \qquad (29)
\end{aligned}$$

since the integral

$$\int_0^{\pi/2} \frac{(\sin\phi)^{n-d_0-1}(\cos\phi)^{n-d_\ell-1}\exp((\sin^2\phi)/2)}{\exp((1/2)\sin^2\phi(RSS_0/RSS_\ell))}\, d\phi \le \int_0^{\pi/2} \exp\left(\frac{\sin^2\phi}{2}\left[1-\frac{RSS_0}{RSS_\ell}\right]\right) d\phi$$

when $n \geq d_0 + 1$ and $n \geq d_\ell + 1$. The latter integral has a finite value for all n according to Casella et al. (2009, p. 1216). Hence, the integral involved in the $BF_{\ell 0}^{\text{J-PEP}}$ has also a finite value for all n.

If we compare any two models M_ℓ and M_k (both of them different than the reference model) we have that

$$-2\log BF_{\ell k}^{\text{J-PEP}} \approx n\log\frac{RSS_\ell}{RSS_k} + (d_\ell - d_k)\log n = BIC_\ell - BIC_k. \qquad (30)$$

Therefore, the J-PEP approach has the same asymptotic behavior as the BIC-based variable-selection procedure. The following lemma is a direct result of (30) and of Theorem 4 of Casella et al. (2009).

Lemma 1. *Let $M_\ell \in \mathcal{M}$ be a normal regression model of type* (2) *such that*

$$\lim_{n\to\infty} \frac{\mathrm{X}_T(\mathrm{I}_n - \mathrm{X}_\ell(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}\mathrm{X}_\ell^T)\mathrm{X}_T}{n} \text{ is a positive semidefinite matrix,}$$

with X_T being the design matrix of the true data generating regression model $M_T \neq M_\ell$. Then the variable selection procedure based on J-PEP Bayes factor is consistent since $BF_{\ell T}^{J\text{-}PEP} \to 0$ as $n \to \infty$.

4 Simulation study

In this section, we perform a simulation comparison that studies the behavior of the proposed method as the sample size increases. We compare the performance of our method with that of the "most established" Bayesian variable selection techniques: the g-prior (Zellner, 1976), the hyper-g prior (Liang et al., 2008), the Zellner and Siow (1980) prior and the BIC (Schwarz, 1978). All competing methods were implemented using the `BAS` package in `R`; we set $g = n$ in the g-prior to correspond to the unit information prior (Kass and Wasserman, 1995) and $\alpha = 3$ in the hyper-g prior as recommended by Liang et al. (2008). For the implementation of our approach, we used the second Monte Carlo scheme presented in Section 3 of Fouskakis, Ntzoufras and Draper (2015).

We consider 100 simulated data-sets of sample sizes $n = 30, 50, 100, 500, 1000$ and $p = 10$ covariates generated from a standardized normal distribution, while the response is generated from

$$Y_i \sim N(0.3X_{i3} + 0.5X_{i4} + X_{i5}, 2.5^2) \qquad \text{for } i = 1, \dots, n. \tag{31}$$

Figure 1 depicts the between-samples distribution of the posterior probability of the true model for the Bayesian variable selection techniques under comparison. It is clear that for small sample sizes all competitive methods fail to provide high posterior evidence in favor of the true model. As the sample size gets larger, all methods increase their posterior support toward the true model, with the proposed J-PEP method to perform slightly better than the Zellner's g-prior and the BIC. This is sensible since these three methods are converging to the same Bayes factors as n grows but with J-PEP constantly supporting more parsimonious models. On the other hand, the hyper-g prior gives the lowest support toward the true model due to its hierarchical structure which increases the posterior uncertainty on the model space. Practically, the hyper-g prior needs larger sample size than the rest of the methods, in order to fully a-posteriori support the true generating mechanism.

Looking now at the posterior inclusion probabilities of each covariate in Figure 2, we observe that all methods successfully identify X_5 (with true effect equal to one) as an important component of the model, even for small sample sizes, with the exception of the Zellner's g-prior. Furthermore, the between-samples variability of the posterior inclusion probabilities reduces as the sample size increases. Returning back to the Zellner's g-prior, it fails to a-posteriori support X_5 for $n = 30$

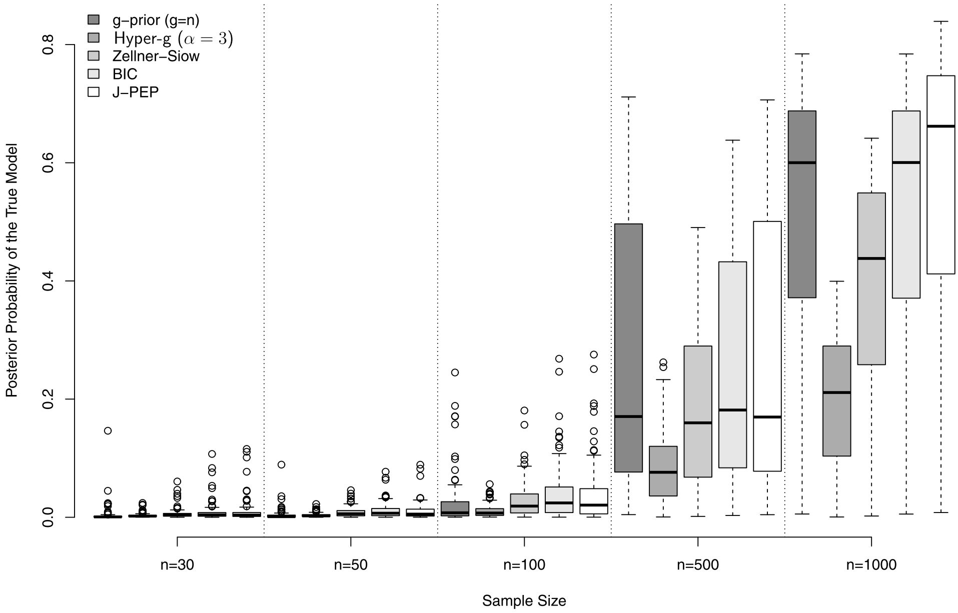
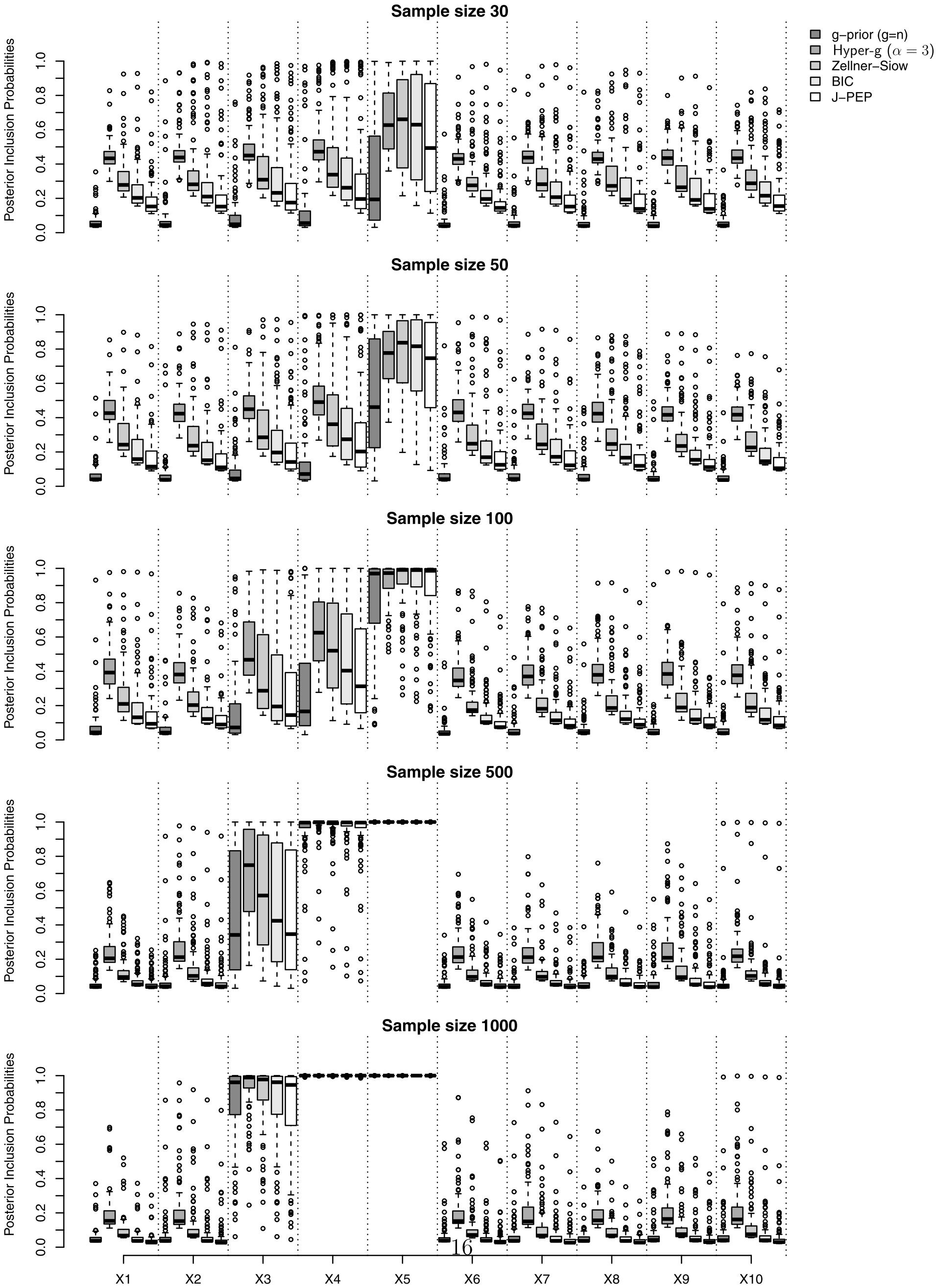


Figure 1 *Boxplots* (*per* 100 *simulated datasets of different sample sizes*) *of the posterior probability of the true model for different variable selection methods.*

and $n = 50$. Generally, the g-prior demonstrates much larger between-sample variability than the rest of the methods and it seems to be unable to identify the true effects for small sample sizes in this simulation study.

Similar is the picture for the posterior inclusion probabilities of the other two covariates with nonzero effects, X_3 and X_4, but with slower rates of convergence toward to one. For the latter covariate (with true effect equal to 0.5), we observe large between-samples uncertainty concerning the importance of this effect for $n \leq 100$ under all methods. For $n \geq 500$, all methods successfully identify the importance of this covariate with small between-samples variability. In general, the hyper-g method supports this covariate with the highest inclusion probabilities while the J-PEP with the lowest inclusion probabilities. This is due to the characteristics of the two methods, with the first supporting more complicated models while the latter more parsimonious ones. We reach to similar conclusions for covariate X_3 (with true effect equal to 0.3) but with the addition that the Zellner's g-prior does not spot the effect of this covariate as important, even for samples of size $n = 500$. Moreover, we need to increase the sample size to $n = 1000$, for all methods, in order to obtain high posterior inclusion probabilities with relatively low between-samples variability.

Reasonably, the between-samples distribution of the posterior inclusion probabilities is similar for all covariates with zero true effects. It is noticeable that all methods, except the hyper-g prior, identify, really fast, that these covariates should



16

Figure 2 *Boxplots (per* 100 *simulated datasets of different sample sizes) of posterior inclusion probabilities for each covariate under the different variable selection methods.*

have low posterior inclusion probabilities with the between-samples variability considerably to decrease as n gets larger. On the other hand, the posterior inclusion probabilities under the hyper-g prior setup are systematically higher (close to 0.5) than the corresponding ones under the other competing methods. This increases the posterior uncertainty on the model space and results to lower probabilities of identifying the true model as the maximum a-posteriori model. It is also noticeable that these posterior inclusion probabilities, under the hyper-g prior setup, both in terms of median values and in terms of between-samples variability, seem to converge very slowly toward zero as n gets larger.

To sum up, in this simulation study the J-PEP prior methodology identifies the true model structure with (slightly) higher posterior probability than the rest of the methods. It provides posterior inclusion probabilities close to zero for nonimportant effects (even for small sample sizes) and high inclusion probabilities for the important effects (although these are smaller than the ones obtained under the competing methods for small sample sizes).

5 Discussion

Under the *power-expected-posterior prior* (PEP) approach, ideas from the power-prior and unit-information-prior methodologies are combined. As a result, the PEP priors are minimally-informative and the effect of training samples is reduced. When using the independence Jeffreys as a baseline prior for normal linear models, we prove that the J-PEP approach has the same asymptotic behavior as the BIC-based variable-selection procedure. Therefore, under very mild conditions on the design matrix, it is a consistent variable selection technique.

Acknowledgments

We wish to thank the Editor and the referee for comments that greatly strengthened the paper. This research has been co-financed in part by the European Union (European Social Fund-ESF) and by Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)-Research Funding Program: Aristeia II/PEP-BVS.

References

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic Bayes factor for linear models. In *Bayesian Statistics* (J. Bernardo, J. Berger, A. Dawid and A. Smith, eds.) **5**, 25–44. New York: Oxford Univ. Press. MR1425398

Casella, G., Girón, F. J., Martínez, M. L. and Moreno, E. (2009). Consistency of Bayesian procedures for variable selection. *The Annals of Statistics* **37**, 1207–1228. MR2509072

Fouskakis, D., Ntzoufras, I. and Draper, D. (2015). Power-expected-posterior priors for variable selection in Gaussian linear models. *Bayesian Analysis* **10**, 75–107. MR3420898

Good, I. J. (2004). *Probability and the Weighting of Evidence*. New York: Haffner.

Ibrahim, J. G. and Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science* **15**, 46–60. MR1842236

Iwaki, K. (1997). Posterior expected marginal likelihood for testing hypotheses. *Journal of Economics, Asia University* **21**, 105–134.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**, 928–934. MR1354008

Liang, F., Paulo, R., Molina, G., Clyde, M. A. and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association* **103**, 410–423. MR2420243

O'Hagan, A. (1995). Fractional Bayes factors for model comparison. *Journal of the Royal Statistical Society, Series B* **57**, 99–138. MR1325379

Pérez, J. M. and Berger, J. O. (2002). Expected-posterior prior distributions for model selection. *Biometrika* **89**, 491–511. MR1929158

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464. MR0468014

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. *Statistics in Practice*. Chichester: Wiley.

Spiegelhalter, D. J. and Smith, A. (1988). Bayes factors for linear and log-linear models with vague prior information. *Journal of the Royal Statistical Society, Series B* **44**, 377–387. MR0693237

Zellner, A. (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms. *Journal of the American Statistical Association* **71**, 400–405. MR0405699

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypothesis (with discussion). In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.) **1**, 585–606 & 618–647 (discussion). Oxford: Oxford Univ. Press. MR0638871

Department of Mathematics
National Technical University of Athens
Zografou Campus
Athens 15780
Greece
E-mail: fouskakis@math.ntua.gr

Department of Statistics
Athens University of Economics and Business
76 Patision Street
Athens 10434
Greece
E-mail: ntzoufras@aueb.gr