UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

JOSE ALEJANDRO ORDOÑEZ CUASTUMAL

On default priors for regression analysis

Distribuições a priori padrão em análise de regressão

Campinas

2021

Jose Alejandro Ordoñez Cuastumal

On default priors for regression analysis

Distribuições a priori padrão em análise de regressão

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Estatística.

Thesis presented to the Institute of Mathematics, Statistics and Scientific Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Statistics.

Supervisor: Victor Hugo Lachos Dávila

Co-supervisor: Larissa Avila Matos

Este trabalho corresponde à versão final da Tese defendida pelo aluno Jose Alejandro Ordoñez Cuastumal e orientada pelo Prof. Dr. Victor Hugo Lachos Dávila.

Campinas

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

Or2o

Ordoñez Cuastumal, Jose Alejandro, 1993-
On default priors for regression analysis / Jose Alejandro Ordoñez Cuastumal. – Campinas, SP : [s.n.], 2021.

Orientador: Víctor Hugo Lachos Dávila.
Coorientador: Larissa Avila Matos.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Geologia - Métodos estatísticos. 2. Student-t multivariada. 3. Inferência bayesiana. 4. Estimativas a-priori. 5. Análise de regressão. I. Lachos Dávila, Víctor Hugo, 1973-. II. Matos, Larissa Avila, 1987-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Distribuições a priori padrão em análise de regressão
Palavras-chave em inglês:
Geology - Statistical methods
Multivariate Student-t
Bayesian inference
A priori estimates
Regression analysis
Área de concentração: Estatística
Titulação: Doutor em Estatística
Banca examinadora:
Víctor Hugo Lachos Dávila [Orientador]
Verónica Andrea González-López
Luis Mauricio Castro Cepero
Aldo William Medina Garay
Pedro Luiz Ramos
Data de defesa: 26-04-2021
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-8157-9012
- Currículo Lattes do autor: http://lattes.cnpq.br/7635604449274955

Tese de Doutorado defendida em 26 de abril de 2021 e aprovada

pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). VICTOR HUGO LACHOS DÁVILA

Prof(a). Dr(a). VERÓNICA ANDREA GONZÁLEZ LÓPEZ

Prof(a). Dr(a). LUIS MAURICIO CASTRO CEPERO

Prof(a). Dr(a). ALDO WILLIAM MEDINA GARAY

Prof(a). Dr(a). PEDRO LUIZ RAMOS

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Acknowledgements

First, I would like to thank god for accompanying me throughout my career, for being my strength in my weakness moments and for give me a life full of learning and experiences.

I thank my parents Maria and Jose for supporting me at all times, For the values they have instilled in me and give me the opportunity to have an excellent education throughout my life.

I am also grateful with my brothers John and Sebastian for being an important part of my life and represent the family unit.

I would like to express my gratitude to my advisors, the professors Victor Hugo and Larissa Matos and to the professor Marcos Oliveira Prates for their support during my research process and the production of the papers presented here.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Resumo

Como conhecido na literatura, a escolha da distribuição a priori é um aspecto chave no método Bayesiano e em muitos casos não é trivial. Neste trabalho, duas distribuições a priori padrão foram derivadas para realizar inferência Bayesiana nos modelos Student-t espacial e de regressão binária.

No primeiro caso, sob a perspectiva da análise Bayesiana objetiva, foi introduzida uma priori baseada no método de referência e duas prioris de Jeffreys para o modelo Student-t espacial (T-SR) considerando os graus de liberdade desconhecidos. Além disso, foram mostradas as condições sob as quais estas densidades geram distribuições à posteriori próprias. Estudos de simulações e uma aplicação em dados reais foram usados com o objetivo de avaliar o rendimento das prioris mencionadas, incorporando uma priori vaga no processo de comparação.

Com relação ao modelo de regressão binária, a distribuição a priori com complexidade penalizada foi introduzida para o parâmetro de assimetria na família de funções de ligação potência, úteis para lidar com dados não balanceados. Uma expressão geral para a distribuição a priori foi derivada e a sua utilidade mostrada para alguns casos particulares, como por exemplo as ligações logito e probito potência. Adicionalmente, foi mostrado que a ligação log-log complementar e sua ligação reversa associada são não identificáveis. Para a obtenção de amostras a posteriori o algoritmo de Monte Carlo Hamiltoniano foi implementado, evitando, assim, os comportamentos de caminhada aleatória dos algoritmos de amostragem Metropolis-Hastings e Gibbs. Um estudo de simulação e uma aplicação em dados reais foram usados para avaliar o rendimento das distribuições a priori introduzidas quando comparadas com as densidades Gaussiana e uniforme.

Palavras-chave: Geoestatística, Distribuição Student-t multivariada, Priori objetiva, Distribuição à priori própria, Distribuição à priori Priori de complexidade penalizada, Estatística Bayesiana, Ligações de Potência, Regressão binária,

Abstract

As widely indicated in the literature, the choice of the prior distribution is a key aspect of the Bayesian method, and in many cases is not trivial. In this work, we derive two default prior distributions to perform Bayesian inference using the spatial Student-t and binary regression models.

In the first case, we use the objective Bayesian analysis framework to introduce a *reference*-based and two *Jeffreys* priors for the spatial Student-t regression (T-SR) model with unknown degrees of freedom. We also show the conditions under which these densities yield to a proper posterior distribution. Simulation studies and a real data application are used to evaluate the performance of the mentioned priors, incorporating a *vague* one in the comparison process.

Regarding the binary regression model, we introduce the penalized complexity prior of the skewness parameter α for the family of power links, which are useful to deal with imbalanced data. We derive a general expression for this density and show its usefulness in some particular cases, such as the power logit and the power probit links. In addition, we show that the power complementary log-log and its associated reversal link are non-identifiable. To obtain posterior samples, we use Hamiltonian Monte Carlo, which avoids the random walk behavior of the Metropolis and Gibbs sampling algorithms. A simulation study and a real data application are used to assess the efficiency of the introduced densities in comparison with the Gaussian and uniform densities.

Keywords: Geostatistics, Multivariate Student-t distribution, Objective prior, Proper prior, PC prior, Bayesian statistics, Power links, Binary regression.

List of Figures

List of Tables

Contents

1 Introduction

The Bayesian method can be used to analyze information in scientific, medical and socioeconomic problems. Its main advantage is to fully account for the parameter uncertainty when performing prediction and inference, even in small samples (Berger *et al.*, 2001). This practice has evolved into an important tool that allows the use of statistical methods jointly with prior knowledge (usually provided by an expert), to investigate a phenomenon of interest (Leonard, 2014). One of the main challenges of Bayesian statistics is the elicitation of prior distributions. There are many circumstances in which either there is no information at all, or there is only a small amount of previous knowledge about the problem being studied. The main purpose of this work is to propose default priors to deal with these situations in the context of regression analysis.

In the first part of this study, we propose a novel method to develop Bayesian inference in the Student-t spatial regression model without the need for prior information. We focus our effort on developing an objective analysis based on the Jeffreys and reference priors. The first one, was proposed by Jeffreys (1946) and is related to the information matrix. One of the main advantages is that these priors are invariant under different reparameterizations, so they work particularly well in the one-dimensional case. Nonetheless, these priors have received several criticisms and various problems have been detected when they are applied to multiparameter scenarios (see, for instance, Syversveen, 1998; Aldrich, 2008). The reference prior, on the other hand, was introduced by Bernardo (1979) as an extension of Jeffreys' work. This prior overcomes many of the inadequacies of Jeffreys' method, by decomposing the parameter space into spaces of lower dimensions. Unfortunately it still has some drawbacks when dealing with hierarchical models because its derivation is challenging. Simpson *et al.* (2017).

Regarding the second part, we work from the perspective of weakly informative priors in the context of regression models applied to binary, binomial and bounded data. The most common practice in this case, is to make use of prior distributions previously/wrongly employed in the literature or to specify new priors based on historical studies (Lemoine, 2019). Another common practice is to choose conjugated priors to reduce both analytical and computational efforts (Robert, 2007). These ad-hoc methods have received criticisms. On one hand, the use of priors from previous research can lead to incorrect inferences, since they were designed specifically to solve a particular problem and probably will not work with adequate efficiency in prediction problems. Moreover, the use of conjugate priors, despite of their analytical advantages, cannot be justified at all, since they trade these facilities for the subjective determination of the prior distribution, which can leads to ignoring part of the prior information. Our aim here is to introduce a new density function for the skewness parameter of power links based on the penalized complexity (PC) prior proposed by (Simpson *et al.*, 2017). Under the assumption of the existence of a baseline model, the PC prior uses the Kullback- Leibler divergence to measure the loss of information when this model is used instead of a more flexible one. This proposal can be weakly informative or informative, depending on the researcher's knowledge about the size of the complexity parameter. The PC prior has desirable properties like invariance under reparameterizations, computational feasibility, and no need to fix prior specifications based on previous research. Nonetheless, the concept of "baseline" model is more subjective than objective. Also, the fact of leaving the specification of the complexity parameter to the user is not very realistic when dealing with models having hierarchical structures.

1.1 Preliminaries

In this section, we give a brief description of some basics settings used in this work, these include models definition, covariance structures and Bayesian and mathematical background.

1.1.1 The Student-t spatial regression Model

Let $Y(\mathbf{s}_i)$ denote the response over location $\mathbf{s}_i \in D_\mathbf{s}$ where $D_\mathbf{s}$ is a continuous spatial domain in $\mathbb{R}^2$. We assume that the observed data $\mathbf{y} = (y(\mathbf{s}_1), \ldots, y(\mathbf{s}_n))^\top$ is a single realization of a Student-t stochastic process $Y(\mathbf{s}) \equiv y(\mathbf{s}) : \mathbf{s} \in D_\mathbf{s}$ (Palacios & Steel, 2006). Thus, if $\mathbf{Y}$ follows a multivariate Student-t distribution with location vector μ, scale matrix Σ and ν degrees of freedom, $\mathbf{Y} \sim t_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}, \nu)$. A Student-t spatial regression (T-SR) model can be represented as

$$Y(\mathbf{s}_i) = X(\mathbf{s}_i)\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\mathbf{s}_i}, \quad \mathbf{s} = \mathbf{s}_1, \ldots, \mathbf{s}_n, \tag{1.1}$$

where $\mathbf{X}(\mathbf{s}_i)$ is a $n \times p$ nonstochastic matrix of full rank and $\boldsymbol{\epsilon}$ is a $n \times 1$ random vector such that $E[\boldsymbol{\epsilon}] = \mathbf{0}$, which represents the error of the process. Therefore, a realization of a Student-t process can be represented by setting $\boldsymbol{\epsilon} \sim t_n(\mathbf{0}, \boldsymbol{\Sigma}, \nu)$, or equivalently, $\mathbf{Y} \sim t_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}, \nu)$ with a valid covariance function for the scale matrix $\boldsymbol{\Sigma}$. We concentrate on a particular parametric class of covariance functions such that the scale matrix is given by

$$\boldsymbol{\Sigma} = [C(\mathbf{s}_i, \mathbf{s}_j)] = \sigma^2\mathbf{R}(\phi) + \tau\mathbf{I},$$

where in standard geostatistical terms: σ^2 is the sill; τ is the nugget effect; ϕ determines the range of the spatial process; $\mathbf{R}(\phi)$ is an $n \times n$ correlation matrix; and $\mathbf{I}$ is the identity matri of order n. We assume that $\mathbf{R}(\phi)$ is an isotropic correlation matrix and depends only the Euclidean distance $d_{ij} = ||\mathbf{s}_i - \mathbf{s}_j||$ between the points $\mathbf{s}_i$ and $\mathbf{s}_j$. Thus, the likelihood function of the model parameters $\theta = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\phi}, \tau, \nu)^\top$, based on the observed data $\mathbf{y}$, is given by

$$L(\theta|\mathbf{y}) = \frac{\boldsymbol{\Gamma}\left(\frac{\nu+n}{2}\right)\nu^{\frac{\nu}{2}}}{\boldsymbol{\Gamma}\left(\frac{\nu}{2}\right)\pi^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\left\{\nu + (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right\}^{-\frac{\nu+n}{2}}, \tag{1.2}$$

where $|\mathbf{A}|$ denotes the determinant of the matrix $\mathbf{A}$. Note that this process is stationary since $\mu = \mathbf{X}\boldsymbol{\beta}$ is constant for all $\mathbf{s}$ in $\mathbb{R}^2$ and the isotropy of $\mathbf{R}(\phi)$ makes that the covariance matrix only depend on the vector difference between $\mathbf{s}_i$ and $\mathbf{s}_j$. In this work, we consider the following four general families of isotropic correlation functions in $\mathbb{R}^2$:

1. The *Matérn family*, given by:

$$\mathbf{R}(\phi) = \begin{cases} \dfrac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\dfrac{d_{ij}}{\phi}\right)^{\kappa} K_\kappa(d_{ij}/\phi), & d_{ij} > 0, \\ 1, & d_{ij} = 0, \end{cases} \tag{1.3}$$

   where $\phi > 0$; $K_\kappa(u) = \dfrac{1}{2}\displaystyle\int_0^\infty x^{\kappa-1}e^{-\frac{1}{2}u(x+x^{-1})}dx$ is the modified Bessel function of the third kind of order κ (see, Gradshtejn & Ryzhik, 1965), with $\kappa > 0$ fixed.

2. The *power exponential family*, defined by:

$$\mathbf{R}(\phi) = \begin{cases} \exp\left\{-\left(\dfrac{d_{ij}}{\phi}\right)^{\kappa}\right\}, & d_{ij} > 0, \quad 0 < \kappa \leqslant 2, \\ 1, & d_{ij} = 0. \end{cases}$$

   Like the Matérn family, the power exponential has a scale parameter $\phi > 0$, a shape parameter κ, in this case bounded by $0 < \kappa \leqslant 2$. It generates correlation functions which are monotone decreasing in d_{ij}.

3. The *spherical family*, given by:

$$\mathbf{R}(\phi) = \begin{cases} 1 - \dfrac{3}{2}(d_{ij}/\phi) + \dfrac{1}{2}(d_{ij}/\phi)^3, & 0 \leqslant d_{ij} \leqslant \phi, \\ 0, & d_{ij} > \phi. \end{cases}$$

4. The *Cauchy family*, defined as:

$$\mathbf{R}(\phi) = \begin{cases} \left[1 + \left(\dfrac{d_{ij}}{\phi}\right)^2\right]^{-\kappa}, & d_{ij} > 0, \\ 1, & d_{ij} = 0. \end{cases}$$

Note that for spherical family, one qualitative difference with respect to others families described earlier is that it has a finite range i.e., $\mathbf{R}(\phi) = 0$ for sufficiently large d_{ij}, namely $d_{ij} > \phi$.

1.1.2 Basic concepts from Bayesian inference

By modeling both the observed data $\mathbf{y}$ and the unknown parameters θ as random variables, the Bayesian approach to statistical analysis provides a coherent framework for combining complex data and external knowledge or expert opinion (Banerjee *et al.*, 2014). This allows parameter uncertainty in the predictions, giving more realistic estimates of the prediction variance. In this approach, we observe data from a sampling density $f(\mathbf{y}|\theta)$, assigning to θ a prior density $\pi(\theta)$ (prior knowledge) then, through the Bayes theorem, we can update this probabilities using $\mathbf{y}$ as:

$$p(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\displaystyle\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}. \tag{1.4}$$

One of the main objectives of Bayesian analysis is to summarize (1.4) to perform inferences about functions of θ. Due to the form of the resulting posterior, this process involve integrals that only can be evaluated through numerical approximations. Rejection sampling and Monte Carlo Markov Chains (MCMC) methods could be adequate to deal with this problem, both procedures consist of sampling from the posterior distribution and make inferences about a function of θ based on the obtained sample. In this work we give a brief summary of the Metropolis-Hastings and the Gibbs algorithm. These methods are use to sample from the posterior density resulting from the reference and the Jeffreys prior obtained for the spatial Student-t model. For more details, see for example Metropolis *et al.* (1953), Hastings (1970) and Gilks & Wild (1992).

One of the objectives of this thesis is the development of a new prior for the Student-t spatial regression model, based on the reference method. We dedicated the following subsection to give a brief introduction to this procedure in a general context, i.e, considering $f(x)$ as an arbitrary density function.

1.1.3 The reference prior method

Bernardo (1979) initiated the reference prior approach to develop non-informative priors. This procedure removes the need for ad hoc modifications by partitioning multidimensional $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ into the parameters of interest and the nuisance parameters. This procedure is widely described in Berger & Bernardo J. (1991) considering the following notation:

- The separation of $\boldsymbol{\theta}$ in m groups, represented by:

$$\boldsymbol{\theta}_{(1)} = (\theta_1, \ldots, \theta_{n_1}), \boldsymbol{\theta}_{(2)} = (\theta_{n_1+1}, \ldots, \theta_{n_1+n_2}),$$
$$\ldots, \boldsymbol{\theta}_{(i)} = \left(\theta_{N_{i-1}+1}, \ldots, \theta_{N_i}\right), \boldsymbol{\theta}_{(m)} = \left(\theta_{N_{m-1}+1}, \ldots, \theta_k\right),$$

where $N_j = \sum_{i=1}^{j} n_i$.

- The definition of $\boldsymbol{\theta}_{[j]}$ and $\boldsymbol{\theta}_{[\sim j]}$ as:

$$\boldsymbol{\theta}_{[j]} = \left(\theta_{(1)}, \ldots, \theta_{(j)}\right) = (\theta_1, \ldots, \theta_{N_j}),$$
$$\boldsymbol{\theta}_{[\sim j]} = \left(\theta_{(j+1)}, \ldots, \theta_{(m)}\right) = (\theta_{N_j+1}, \ldots, \theta_k).$$

  with the convention $\boldsymbol{\theta}_{[\sim 0]} = \boldsymbol{\theta}$ and $\boldsymbol{\theta}_{(0)}$ vacuous.

- The general definition of the Kullback - Leibler divergence between two densities on Θ as:

$$KLD(g,h) = \int_{\Theta} g(\theta) \log\left(\frac{g(\theta)}{h(\theta)}\right) d\boldsymbol{\theta} \tag{1.5}$$

- Finally, the definition of the density of $Z_t = (X_1, \ldots, X_t)$, a random variable arising from t independent replications of the original experiment as:

$$p(z_t|\boldsymbol{\theta}) = \prod_{i=1}^{t} f(x_i|\boldsymbol{\theta}). \tag{1.6}$$

Having this definitions, the method is described in four steps as in algorithm 1. The following subsection give a brief description of some sampling procedures widely used in Bayesian statistics, they facilitate the inference approximating moments and quantities of interest by the use of random samples.

1.1.4 Metropolis-Hastings method

The Metropolis-Hastings (M-H) algorithm allows to obtain an approximate sample from an univariate density for which is not easy to carry out this procedure. Let f be the target distribution we are interested in sampling and g the proposal distribution. Following Metropolis *et al.* (1953) and Hastings (1970), the method starts at $t = 0$ drawing at random $\theta^{(0)}$ from g under the condition that $f(x^{(0)}) > 0$. Given $\theta^{(t)}$, the method generates $\theta^{(t+1)}$ as shown in algorithm 2.

The generated chain has the Markov property since $\theta^{(t+1)}$ only depend on $\theta^{(t)}$. Irreducibility and aperiodicity of the same are sufficient conditions to ensure that the obtained sample converges to a unique limiting stationary distribution.

1.1.5 Gibbs sampling

This method allows sampling from a multidimensional density function, As in the M-H algorithm, the goal is to obtain a chain with the Markov property whose distribution converges to the target distribution f. Following Geman & Geman (1984),

Algorithm 1 Obtaining the reference prior for a general model.

1: Choose a nested sequence $\left\{\Theta^{\ell}\right\}$ of compact subsets if theta such that, $\cup_{\ell=1}^{\infty}\Theta^{\ell}=\Theta$
2: Order the coordinates $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)$ and divide them into m groups. Usually is better to have $m=k$, and the order should typically be according inferential importance.
3: For $j=m,m-1,\ldots,1$, iteratively compute the densities $\pi\left(\boldsymbol{\theta}_{\sim(j-1)}\middle|\,\boldsymbol{\theta}_{[j-1]}\right)$, using:

$$\pi_j^{\ell}\left(\boldsymbol{\theta}_{\sim(j-1)}\middle|\,\boldsymbol{\theta}_{[j-1]}\right) \;\propto\; \pi_{j+1}^{\ell}\left(\boldsymbol{\theta}_{\sim(j)}\middle|\,\boldsymbol{\theta}_{[j]}\right)h_j^{\ell}\left(\boldsymbol{\theta}_{\sim(j)}\middle|\,\boldsymbol{\theta}_{[j-1]}\right),$$

where $\pi_{m+1}^{\ell}=1$ and h_j^{ℓ} is computed following the next steps:

(a) Define $p_t(\boldsymbol{\theta}_{(j)}|\boldsymbol{\theta}_{[j-1]})$ by:

$$p_t(\boldsymbol{\theta}_{(j)}|\boldsymbol{\theta}_{[j-1]}) \;\propto\; \exp\left(\int p\left(z_t|\boldsymbol{\theta}_{[j]}\right)\log\left(p(\theta_{(j)}|z_t,\boldsymbol{\theta}_{[j-1]})\right)dz_t\right),$$

where (using $p(\cdot)$ generically to represent the conditional density of the given variables)

$$\begin{aligned} p(z_t|\boldsymbol{\theta}_{[j]}) &= \int p(z_t|\boldsymbol{\theta})\pi_{j+1}^{\ell}\left(\boldsymbol{\theta}[\sim j]|\boldsymbol{\theta}_{[j]}\right)d\boldsymbol{\theta}[\sim j], \\ p(\theta_{(j)}|z_t,\boldsymbol{\theta}_{[j-1]}) &\propto p(z_t|\boldsymbol{\theta}_{[j]})p_t(\boldsymbol{\theta}_{(j)}|\boldsymbol{\theta}_{[j-1]}). \end{aligned}$$

(b) Assuming the limit exists, define:

$$h_j^{\ell}(\boldsymbol{\theta}_{(j)}|\boldsymbol{\theta}_{[j-1]}) \;=\; \lim_{t\to\infty}p_t(\boldsymbol{\theta}_{(j)}|\boldsymbol{\theta}_{[j-1]}).$$

4: Define a reference prior $\pi(\boldsymbol{\theta})$, as any prior for which:

$$E_{\ell}^{X}KLD(\pi_1^{\ell}(\boldsymbol{\theta}|X),\pi(\boldsymbol{\theta}|X))\to 0 \text{ as } \ell\to\infty,$$

where KLD is defined in 1.5 and E_{ℓ}^{X} is the expectation with respect to:

$$p^{\ell}(x) \;=\; \int_{\Theta}f(x|\theta)\pi_1^{\ell}(\boldsymbol{\theta})d\boldsymbol{\theta},$$

(writing $\pi_1^{\ell}(\boldsymbol{\theta})$ for $\pi_1^{\ell}(\boldsymbol{\theta}[\sim 0]|\boldsymbol{\theta}[0])$).

Let $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_p)$ and denote $\boldsymbol{\theta}_{-i}=(\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_p)$. Under the assumption that $\theta_i|\boldsymbol{\theta}_{-i}$ can be easily sampled, the Gibbs sampling procedure can be described as in algorithm 3.

It is possible to combine the Gibbs procedure with a M-H sampling allowing a proposal distribution which varies over the time. Each Gibbs cycle will consist of p Metropolis–Hastings steps. To see this, note that the ith Gibbs step in a cycle effectively proposes the candidate vector $\boldsymbol{\theta}^*=\left(\theta_1^{(t+1)},\theta_2^{(t+2)},\ldots,\theta_{i-1}^{(t+1)},\theta_i^{(*)},\theta_{i+1}^{(t)},\ldots,\theta_p^{(t+1)}\right)$ given the current state of the chain $\left(\theta_1^{(t+1)},\theta_2^{(t+1)},\ldots,\theta_{i-1}^{(t+1)},\theta_i^{(t)},\theta_{i+1}^{(t)},\ldots,\theta_p^{(t+1)}\right)$. Thus, the ith univariate Gibbs update can be viewed as a Metropolis–Hastings step drawing,

Algorithm 2 Obtaining a random sample using the Metropolis-Hastings algorithm.

1: Sample a candidate θ^* from the proposal distribution $g\left(.|\theta^{(t)}\right)$
2: Compute the M-H ratio as:

$$R\left(\theta^{(t)}, \theta^*\right) = \frac{f(\theta^*)g\left(\theta^{(t)}|\theta^*\right)}{f(\theta^{(t)})g\left(\theta^*|\theta^{(t)}\right)}$$

3: Sample a value for $\theta^{(t+1)}$ as:

$$f(x) = \begin{cases} \theta^*, \text{ with probability } \min(R\left(\theta^{(t)}, \theta^*\right), 1) \\ \theta^{(t+1)}, \text{ otherwise} \end{cases}$$

4: Increment t and go to step 1.

Algorithm 3 Obtaining a random sample using Gibbs algorithm.

1: Select the starting values $\boldsymbol{\theta}^{(0)}$ and fix $t = 0$
2: Generate,

$$\begin{aligned} \theta_1^{(t+1)}|\cdot &\sim f(\theta_1|\boldsymbol{\theta}_{-1}^{(t)}) \\ \theta_2^{(t+1)}|\cdot &\sim f(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_p^{(t)}) \\ &\vdots \\ \theta_p^{(t+1)}|\cdot &\sim f(\theta_p|\boldsymbol{\theta}_{-p}^{(t+1)}) \end{aligned}$$

3: Increment t to $t + 1$ and go to step 1.

$$\theta^*|\theta_1^{(t+1)}, \ldots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \theta_{i+1}^{(t)}, \ldots, \theta_p^{(t+1)} \sim g_i\left(\cdot|\theta_1^{(t+1)}, \ldots, \theta_{i-1}^{(t+1)}, \theta_i^{(t)}, \theta_{i+1}^{(t)}, \ldots, \theta_p^{(t+1)}\right)$$

It can be shown that in this case the Metropolis–Hastings ratio equals 1, which means that the candidate is always accepted. The Gibbs sampler should not be applied when the dimensionality of $\boldsymbol{\theta}$ changes (e.g., when moving between models with different numbers of parameters at each iteration of the Gibbs sampler).

1.1.6 Hamiltonian Monte Carlo and NUTS sampler

In general, simple methods such as Gibbs sampling (Geman & Geman, 1984) and random walk Metropolis (Metropolis *et al.*, 1953) requires a long time to converge to the target distribution, this is due to its random walk behavior and sensitivity to correlated parameters. An alternative to deal with this problem is the hybrid Monte Carlo (HMC) sampler (Duane *et al.*, 1987; Neal, 2011) which avoids unnecessary iterations by taking a series of steps informed by first-order gradient information. These settings allow it to converge to high-dimensional target distributions much more quickly than other common methods. The algorithm uses Hamiltonian dynamics to simulate a random process via the Störmer-Verlet ("leapfrog") integrator,

$$\mathbf{r}^{t+\frac{\epsilon}{2}} = \mathbf{r}^t + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}^t|\mathbf{y}), \quad \boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t + \epsilon\mathbf{r}^{t+\frac{\epsilon}{2}}, \quad \mathbf{r}^{t+\epsilon} = \mathbf{r}^{t+\frac{\epsilon}{2}} + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}^{t+\epsilon}|\mathbf{y}),$$

where $\mathbf{r}^t$ is an auxiliary momentum variable at time t and $\nabla_{\boldsymbol{\theta}}$ denotes the gradient with respect to $\boldsymbol{\theta}$. Obtaining the momentum variables using this sampler, L leapfrog updates are applied to the position and the momentum variables, generating a proposal position-momentum pair. The proposal $(\tilde{\mathbf{r}}, \tilde{\boldsymbol{\theta}})$ is accepted following a Metropolis reject step with probability,

$$\min\left\{1, \frac{\exp\left(\ell(\tilde{\boldsymbol{\theta}}|\mathbf{y}) - \frac{1}{2}\tilde{\mathbf{r}}^\top\tilde{\mathbf{r}}\right)}{\exp\left(\ell(\boldsymbol{\theta}^{t-1}|\mathbf{y}) - \frac{1}{2}\mathbf{r}^{0\top}\mathbf{r}^0\right)}\right\},$$

where r^0 are the sampled momentum variables before they are put through the leapfrog integrator. Despite the advantages of this method over simpler alternatives, the requirement of the gradient and the user's specification of the step size ϵ and the number of steps L represents a challenge to its use. An extension of the HMC algorithm called NUTS (Non-U Turn Sampler) (Hoffman & Gelman, 2014) avoids these issues by setting ϵ and L automatically and computing the gradient ∇_θ through automatic differentiation.

1.1.7 Big O notation

Let f be a real or complex valued function and g a function defined in the positive real numbers. Following Landau (1909), we say $f(x) = O(g(x))$ if there is a real number $C > 0$ and a real number x_0 such that

$$|f(x)| \leqslant Cg(x), \qquad \text{for all } x \geqslant x_0.$$

This notation is used to describe the behavior of certain functions at limiting points. The Big-O notation is useful in this work, since we are interest in knowing the behavior of some functions at the extremes of their domain, in this way, we are able to establish boundaries which are helpful in showing propriety for the prior distributions we are studying. Some of the most important properties of this concept are,

- If a is a constant, then
$$af = O(f).$$
- If $f_1 = O(g_1)$ and $f_2 = O(g_2)$, then
$$f_1 + f_2 = O(max(g_1, g_2)).$$

- If $f_1 = O(g_1)$ and $f_2 = O(g_2)$, then

$$f_1 f_2 = O(g_1 g_2).$$

1.1.8 Generalized linear models.

Following McCullagh & Nelder (1989), the Generalized Linear Model (GLM) is a flexible extension of ordinary linear regression which allows the use of response variables with errors distributed different from the normal distribution. This generalization has three fundamental characteristics,

1. **The random component:** In this case the vector of responses $\mathbf{Y}$, which were established to belong to the exponential family.
2. **The linear predictor:** Usually denoted by $\eta = \mathbf{X}\boldsymbol{\beta}$.
3. **The link function:** Which is a function g which related the expected value of $\mathbf{Y}$ given a set of covariates with the linear predictor through the equation,

$$g^{-1}(\mu) = \eta.$$

In this thesis, we work with the GLM particular case where $\mathbf{Y}$ follows a binomial or a bounded distribution, i.e, the binary regression model. The following subsection gives a brief description of its components including the family of links we consider.

1.1.8.1 Bounded regression

Consider $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ an $n \times 1$ vector of n independent random variables representing either a binomial or a bounded (in $(0, 1)$) response of n individuals. Also, let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ be the vector of covariates associated with individual i. Then, the hierarchical regression model for these types of responses is of the form:

$$\begin{aligned} Y_i &\sim f(y_i|\mu_i, \theta), \\ g(\mu_i) &= F^{-1}(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}, \end{aligned} \tag{1.7}$$

where f denotes the probability density function (pdf) of each Y_i, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ are the fixed effects, $g = F^{-1}()$ is typically called the link function in the context of generalized linear models, where F is a the cumulative distribution function (cdf) of a random variable with support in the real line. When F is symmetric, then the resulting link is called a symmetric link, while if F is asymmetric, g is called a skewed link. The logit and probit functions are the most popular specifications for g when analyzing these kinds of responses. They are symmetric and obtained from the cdf of the logistic and the standard normal distributions, respectively. The complementary log-log is a popular skewed one, obtained by taking F as the reversal Gumbel distribution.

In this work, we define a PC prior for the asymmetry parameter α of the power link family (Bazán *et al.*, 2017; Lemonte, 2017) whose cumulative cdf and pdf are, respectively, of the form,

$$
\begin{aligned}
F_{\boldsymbol{\theta}}(x) &= F(x|\boldsymbol{\theta}) = \left[F\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha}, \\
f_{\boldsymbol{\theta}}(x) &= f(x|\boldsymbol{\theta}) = \frac{\alpha}{\sigma}\left[F\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1} f\left(\frac{x-\mu}{\sigma}\right),
\end{aligned} \tag{1.8}
$$

where $\boldsymbol{\theta} = (\mu, \sigma, \alpha)$, $F(\cdot)$ denotes any absolute continuous cdf of a random variable with support in the real line and $f(\cdot)$ denotes its pdf which must be unimodal and log-concave.

1.1.9 Model selection criteria

In this section we give a brief description of some of the most used information criteria in the literature.

The deviance information criteria (DIC) is a generalization of Akaike's Information criteria (AIC) given by,

$$DIC = \overline{D}(\boldsymbol{\theta}) + \rho_D = 2\overline{D}(\boldsymbol{\theta}) - D(\tilde{\boldsymbol{\theta}}),$$

where $\tilde{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{y})$, $D(\boldsymbol{\theta}) = -2E(\log f(\mathbf{y}|\boldsymbol{\theta})|\mathbf{y})$ is the deviance of the considered model, $\overline{D}(\boldsymbol{\theta})$ is the posterior expectation of $D(\boldsymbol{\theta})$ (in this case a random variable) and ρ_D defined as $D(\boldsymbol{\theta}) - D(\tilde{\boldsymbol{\theta}})$ is a measure of the effective number of parameters in the model. In general, the computation of $\overline{D}(\boldsymbol{\theta})$ is quite complex, a good approximation can be easily obtained from the MCMC sample $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$ using the sample mean of the deviations $\overline{D} = -2\frac{1}{M}\sum_{m=1}^{M} \log(f(\mathbf{y}|\boldsymbol{\theta}_m)|\mathbf{y})$, in this way we get the estimator $\widehat{DIC} = \overline{D} - D(\tilde{\boldsymbol{\theta}})$.

The expected Akaike information (EAIC) and expected Bayesian information (EBIC) criteria discussed in Brooks (2002) are given by,

$$EAIC = \overline{D} + 2p \quad EBIC = \overline{D} + p\log(n),$$

with p the number of model parameters and n, the effective sample size. Additionally, we use a recent proposal of Watanabe (2010), the widely applicable information criteria (WAIC) which can be seen as approximation to cross-validation in Ando (2011) with an only difference in the computation of the parameter of complexity ρ_{WAIC},

$$\rho_{WAIC} = \sum_{i=1}^{n} V_{m=1}^{M}(\log f(y_i|\boldsymbol{\theta}_m)),$$

with $V_{m=1}^{M}(a) = {}^{1}\!/_{M-1} \sum_{m=1}^{M}(a_m - \bar{a})^2$ and $\bar{a} = {}^{1}\!/_{M} \sum_{m=1}^{M} a_m$. Then,

$$WAIC = \overline{D} + 2\rho_{WAIC}.$$

Prediction measures in binary regression

An useful measure to assess the quality of the estimated rates in a Binomial regression is the square root of the mean square of error (MSE), this is given by,

$$\sqrt{MSE} = \sqrt{\sum_{i=1}^{n_{bin}} \frac{(\hat{p}_i - p_i)^2}{n_{bin}}},$$

where $\hat{p}_i$ is the fitted proportion obtained from the model of interest, p_i is the observed sample proportion and n_{bin} is the number of binomial observations.

The use of the confusion matrix is very common to perform classification in binary regression, its usual structure is showed in Table 1. The response 1 indicates an observed or predicted success, and 0 represents an observed or predicted failure. The terms of the diagonal are the number of times in which the model makes correct predictions, i.e, True Positives (TP) and True Negatives (TN), the remaining terms indicates the frequency in which the model makes incorrect predictions, that is, false positives (FP) and false negatives (FN).

Table 1 – General structure of the confusion matrix.

	Predicted	
Observed	0	1
0	TN	FP
1	FN	TP

As in Fawcett (2006), we use the most common measures for classification, accuracy (ACC), sensitivity or true positive rate (TPR), specificity or true negative rate (TNR) and area under the ROC curve (AUC), these are given by,

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \quad TPR = \frac{TP}{TP + FP}, \quad TNR = \frac{TN}{TN + FN}.$$

Also, given the unbalanced nature of data, we use the critical success index (CSI), Sokal & Sneath index (SSI), Faith index (FAITH) and distance measure pattern difference (PDIF) as were proposed in de la Cruz *et al.* (2019) to measure the similarity between the observed and predicted classification. These measures are defined by,

$$CSI = \frac{TP}{TP + FP + FN} \qquad SSI = \frac{TP}{TP + 2(FP + FN)}$$
$$FAITH = \frac{TP + 0.5TN}{TP + TN + FP + FN} \quad PDIF = \frac{4FPFN}{(TP + TN + FP + FN)^2}.$$

Excepting the PDIF criterion, for all criteria mentioned above, the smaller the values, the better the model fit.

1.1.10 Randomized normalized quantile residuals

To verify the adequacy of the model described by equations (1.7) - (1.8), we can use the normalized randomized quantile residuals defined in Dunn & Smyth (1996). These are given by,

$$r_{q,i} = \Phi^{-1}(u_i),$$

where for binary regression models, u_i is a random value of a uniform distribution in the following interval,

$$\left[I_{1-\hat{\mu}_i}\left(2-y_i, y_i\right), I_{1-\hat{\mu}_i}\left(1-y_i, 1+y_i\right)\right],$$

with $I_x(a,b) = {}^{B(x;a,b)}/_{B(a,b)}$ being the regularized incomplete beta function and $\hat{\mu}_i = F(x_i^\top \boldsymbol{\beta})$ is the estimated probability through the use of a link function. If the model fits correctly, then the normalized randomized quantile residual has a standard normal distribution, which can be evaluated through an envelope graph as suggested by Atkinson (1985).

1.2 Organization of the thesis

This thesis is composed of six chapters containing default procedures to perform Bayesian estimation for both the Student-t spatial and the bounded regression models. This work is organized as follows.

In Chapters 2 and 3 we provide an objective Bayesian method to analyze the Student-t spatial linear regression model. Specifically, in Chapter 2 we establish theoretical foundations of this analysis, starting by giving a general form of the marginal posterior for the range parameter and the degrees of freedom. After that, we give equations which describe the proposed reference based prior, and the Jeffreys priors, the most common non-informative densities in Bayesian statistics. Finally, we show conditions under which these priors generate proper posteriors. In Chapter 3 we also assess the performance of these methods using simulation studies and a real dataset.

In Chapters 4 and 5, we provide a procedure based on the PC prior to estimate the asymmetry parameter for the family of power links proposed by Lemonte (2017). Similar to the previous case, in Chapter 4 we also describe the theoretical background of this method, starting with basic settings, establishing a general expression for the PC prior for the family of power links, and using this formula in some particular cases such as the power logit and power probit links. Chapter 5, on the other hand, describes a simulation study and a real application to assess the relevance of the proposed procedure. Finally, some concluded remarks are discussed in Chapter 6.

2 Objective Bayesian analysis for geostatistical Student-t processes: theoretical settings

2.1 Introduction

Geostatistical data modeling has now virtually permeated all areas of epidemiology, hydrology, agriculture, environmental science, demographic studies, just to name a few. Here, the prime objective is to account for the spatial correlation among observations collected at various locations, and also to predict the values of interest for non-sampled sites (Cressie, 1993). In this chapter, we will focus on a fully Bayesian approach to analyze spatial data, whose main advantage is that parameter uncertainty is fully accounted for when performing prediction and inference, even in small samples (Berger *et al.*, 2001). However, the elicitation of priors for the correlation parameter in Gaussian processes is a nontrivial task (Kennedy & O'Hagan, 2001).

The problem of inference and prediction for spatial data with Gaussian processes using objective priors has received attention in the recent literature. It started with Berger *et al.* (2001) that develop an exact non-informative prior for unknown parameters of Gaussian random fields by using exact marginalization in the reference prior algorithm (see, Berger & Bernardo J., 1991). Further, Paulo (2005) and Ren *et al.* (2013) generalized the previous results to an arbitrary number of parameters in the correlation matrix. After the precursor proposal of De Oliveira (2007), that allows the inclusion of measurement error for reference prior elicitation, other extensions from this perspective were proposed in the literature, see, for instance, Ren *et al.* (2012) and Kazianka & Pilz (2012).

In the context of the Student-t distribution, Zellner (1976) was the first to present a Bayesian and non-Bayesian analysis of a linear multiple regression model with Student-t errors, assuming a scalar dispersion matrix and known degrees of freedom. An interesting result of this work is that inferences about the scale parameter of the multivariate-t distribution can be made using an F-distribution rather than the usual χ^2(or inverted χ^2) distribution. Later, Fonseca *et al.* (2008) developed an objective Bayesian analyses based on the Jeffreys-rule prior and the independence Jeffreys prior for linear regression models with independent Student-t errors and unknown degrees of freedom. This procedure allowed a non-subjective statistical analysis with adaptive robustness to outliers and with a full account of the uncertainty. Branco *et al.* (2013) introduced an objective prior for the shape parameter using the skew-t distribution proposed by Azzalini

& Capitanio (2003). Villa & Walker (2014) constructed an objective prior for the degrees of freedom of the univariate Student-t distribution when this parameter is taken to be discrete. More recently, He *et al.* (2020) proposed objective priors for univariate Student-t regression and study its frequentist properties.

Even though some solutions have been proposed in the literature to deal with the problem of objective prior under the Student-t distribution, to the best of our knowledge there are no studies conducting objective Bayesian analyzes under the Student-t spatial regression model. We introduce a *reference* prior based on the exact marginalization and we derive the conditions that it yields a valid posterior distribution. Moreover, the independence Jeffreys and the Jeffreys-rule priors are derived and analyzed. We show that the Jeffreys priors suffer many drawbacks while the proposed *reference* prior produces more accurate estimates with good frequentist properties.

The chapter is organized as follows. In Section 2.2, a general form of improper priors is presented and the *reference* prior is provided with the conditions of its validity. Section 2.3 introduces the independence Jeffreys and the Jeffreys-rule priors with the necessary conditions to obtain a valid posterior distribution. In Section 2.4.2, model selection criteria are presented in order to evaluate the competing Bayesian models. Technical derivations are relegated to Appendix A.

2.2 The Reference Prior

Palacios & Steel (2006) discuss that the derivation of a reference prior for non-Gaussian processes is not trivial. In this section, we introduce a *reference* prior for the T-SR model defined in (1.1) without a nugget effect, i.e., $\mathbf{\Sigma} = \sigma^2\mathbf{R}(\phi)$. We obtain $\pi(\phi,\nu)$ through the marginal model defined via integrated likelihood.

2.2.1 Prior density for θ

For $\theta = (\boldsymbol{\beta}, \sigma^2, \phi, \nu) \in \Omega = \mathbb{R}^p \times (0,+\infty) \times (0,+\infty) \times (0,+\infty)$, consider the family of improper priors of the form

$$\pi(\theta) \propto \frac{\pi(\phi,\nu)}{(\sigma^2)^a}, \tag{2.1}$$

for different choices of $\pi(\phi,\nu)$ and a. Selection of the prior distribution for ϕ and ν is not straightforward. Assuming an independence structure $\pi(\phi,\nu) = \pi_1(\phi) \times \pi_2(\nu)$, one alternative could be to select improper priors for these two parameters, nevertheless, it is necessary to be careful, since it is obligatory to show that such selection produces a proper posterior distribution. For ϕ, the use of truncation over the parameter space or vague proper priors are alternatives to overcome the improper posterior distribution problem, however, in both cases inferences are often highly dependent on the bounds used or on the hypeparameters selected for the vague distribution (Berger *et al.*, 2001). And, for ν even when the parameter space is restricted, the maximum likelihood estimator may not exist with positive probability (Fonseca *et al.*, 2008; He *et al.*, 2020). Other choices of priors can be found in Geweke (1993); De Oliveira (2007); Ren *et al.* (2012); Kazianka & Pilz (2012); Ren *et al.* (2013); Fonseca *et al.* (2008); Branco *et al.* (2013), but none of this authors consider a Student-t spatial regression framework.

The expression for $\pi(\phi, \nu|\mathbf{y})$ based on an arbitrary prior $\pi(\phi, \nu)$ is presented in the following proposition.

Proposition 2.1. *Consider the model described by (1.1) and (1.2). For $a < \nu/2 + 1$, $\nu > 4$ and different choices for $\pi(\phi, \nu)$, the posterior density $\pi(\phi, \nu|\mathbf{y})$ can be written as*

$$\pi(\phi, \nu|\mathbf{y}) \propto A(\nu) \left|\mathbf{R}\right|^{-\frac{1}{2}} \left|V_{\hat{\beta}}\right|^{\frac{1}{2}} \left\{(n-p)S^2\right\}^{-\left(\frac{n-p}{2}+a-1\right)} \pi(\phi, \nu), \tag{2.2}$$

where $A(\nu) = \nu^{-(1-a)}\Gamma\left(\frac{\nu}{2} - a + 1\right)/\Gamma\left(\frac{\nu}{2}\right)$, $S^2 = \mathbf{z}^\top\mathbf{R}^{-1}\mathbf{z}/(n-p)$, $\mathbf{z} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$ and $V_{\hat{\beta}} = (X^\top \mathbf{R}^{-1} X)^{-1}$.

Let $\hat{\boldsymbol{\beta}}(\phi) = (\mathbf{X}^\top\mathbf{R}^{-1}X)^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{y}$ be the generalized least square estimator of $\boldsymbol{\beta}$. So, given $\pi(\phi, \nu)$, to guarantee propriety of the posterior density $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \nu|\mathbf{y})$ and the existence of the first two moments of the Student-t distribution, we have to ensure that,

$$\int_0^{+\infty}\int_4^{+\infty} A(\nu) \left|\mathbf{R}\right|^{-\frac{1}{2}} \left|V_{\hat{\beta}}\right|^{\frac{1}{2}} \left\{(n-p)S^2\right\}^{-\left(\frac{n-p}{2}+a-1\right)} \pi(\phi, \nu) d\nu d\phi < +\infty. \tag{2.3}$$

Proof. Let $g(\nu) = \Gamma(\nu+n/2)/\nu^{n/2}\Gamma(\nu/2)$, $\nu_1 = n_p + \nu$, $n_p = n - p$, with $p = rank(\mathbf{X})$ and $\nu_2 = \nu_1 + n_p$. To obtain $\pi(\phi, \nu|\mathbf{y})$, first note that

$$\begin{aligned}
\pi(\phi, \nu|\mathbf{y}) &\propto \int_0^{+\infty}\int_{\mathrm{IR}^p} L(\boldsymbol{\beta}, \sigma^2, \phi, \nu)\pi(\boldsymbol{\beta}, \sigma^2, \phi, \nu) d\boldsymbol{\beta} d\sigma^2 \\
&\propto \int_0^{+\infty}\int_{\mathrm{IR}^p} \frac{g(\nu)\nu^{\frac{(\nu+n)}{2}}}{(\sigma^2)^{\frac{n}{2}}\left|\mathbf{R}\right|^{\frac{1}{2}}} \left\{\nu + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma^2}\right\}^{-\frac{n+\nu}{2}} \\
&\quad \frac{\pi(\phi, \nu)}{(\sigma^2)^a} d\boldsymbol{\beta} d\sigma^2 \\
&\propto \frac{\nu^{\frac{(\nu+n)}{2}} g(\nu)\pi(\phi, \nu)}{\left|\mathbf{R}\right|^{\frac{1}{2}}} \\
&\quad \int_0^{+\infty}\int_{\mathrm{IR}^p} \frac{1}{(\sigma^2)^{\frac{n}{2}+a}} \left\{\nu + \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^\top\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}{\sigma^2}\right\}^{-\frac{n+\nu}{2}} d\boldsymbol{\beta} d\sigma^2.
\end{aligned}$$

Now, let $\mathbf{V}_{\hat{\beta}} = \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X}$, $\tilde{\sigma}^2 = \dfrac{\sigma^2\nu + n_p S^2}{\nu_1}$ and $S^2 = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})^\top\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}})/(n-p)$, with $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{y}$. After some algebra manipulations, we have that

$$\pi(\phi, \nu|\mathbf{y}) \propto \int_0^{+\infty} \frac{(\sigma^2)^{\frac{\nu}{2}-a}}{(\tilde{\sigma}^2)^{\frac{n+\nu}{2}}} \int_{\mathrm{IR}^p} \left\{\nu_1 + \frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top\mathbf{V}_{\hat{\beta}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{\tilde{\sigma}^2}\right\}^{-\frac{\nu_1+p}{2}} d\boldsymbol{\beta} d\sigma^2$$

$$\frac{\nu^{\frac{(\nu+n)}{2}} g(\nu)\pi(\phi,\nu)}{|\mathbf{R}|^{\frac{1}{2}}}.$$

Note that, the term inside of the integral with respect to $\boldsymbol{\beta}$ is the kernel of the $t_p(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\hat{\beta}}, \nu_1)$ distribution. Hence,

$$\pi(\phi,\nu|\mathbf{y}) \propto \frac{\nu^{\frac{(\nu+n)}{2}} g(\nu) \left|\mathbf{V}_{\hat{\beta}}\right|^{\frac{1}{2}}}{\nu_1^{\frac{\nu_1+p}{2}} g(\nu_1)\, |\mathbf{R}|^{\frac{1}{2}}} \pi(\phi,\nu) \int_0^{+\infty} \frac{(\sigma^2)^{\frac{\nu}{2}-a}}{(\tilde{\sigma}^2)^{\frac{\nu_1}{2}}} d\sigma^2 \propto c_1 \int_0^{+\infty} \frac{(\sigma^2)^{\frac{\nu}{2}-a}}{(\tilde{\sigma}^2)^{\frac{\nu_1}{2}}} d\sigma^2,$$

where $c_1 = \dfrac{\nu^{\frac{(\nu+n)}{2}} g(\nu) \left|\mathbf{V}_{\hat{\beta}}\right|^{\frac{1}{2}}}{\nu_1^{\frac{\nu_1+p}{2}} g(\nu_1)\, |\mathbf{R}|^{\frac{1}{2}}} \pi(\phi,\nu)$ and $g(\nu_1) = \Gamma(\nu_1+p/2)/\nu_1^{p/2}\Gamma(\nu_1/2)$. Now,

$$\begin{aligned}
\int_0^{+\infty} \frac{(\sigma^2)^{\frac{\nu}{2}-a}}{(\tilde{\sigma}^2)^{\frac{\nu_1}{2}}} d\sigma^2 &\propto \int_0^{+\infty} (\sigma^2)^{\frac{\nu}{2}-a} \left(\frac{\sigma^2\nu + n_p S^2}{\nu_1}\right)^{-\frac{\nu_1}{2}} d\sigma^2 \\
&\propto \int_0^{+\infty} \frac{\nu_1^{\frac{\nu_1}{2}} (\sigma^2)^{\frac{\nu}{2}-a}}{(n_p S^2)^{\frac{\nu_1}{2}}} \left(1 + \frac{\nu}{n_p}\frac{\sigma^2}{S^2}\right)^{-\frac{\nu_1}{2}} d\sigma^2 \\
&\propto \frac{\nu_1^{\frac{\nu_1}{2}} (S^2)^{-(\frac{n_p}{2}+a)}}{n_p^{\frac{\nu_1}{2}}} \int_0^{+\infty} \left(\frac{\sigma^2}{S^2}\right)^{\frac{\nu}{2}-a} \left(1 + \frac{\nu}{n_p}\frac{\sigma^2}{S^2}\right)^{-\frac{\nu_1}{2}} d\sigma^2 \\
&\propto \frac{\nu_1^{\frac{\nu_1}{2}} (S^2)^{-(\frac{n_p}{2}+a)}}{n_p^{\frac{\nu_1}{2}}} \\
&\quad \int_0^{+\infty} \left(\frac{\sigma^2}{S^2}\right)^{1-a} \left(\frac{\sigma^2}{S^2}\right)^{\frac{\nu}{2}-1} \left(1 + \frac{\nu}{n_p}\frac{\sigma^2}{S^2}\right)^{-\frac{\nu_1}{2}} d\sigma^2.
\end{aligned}$$

Let $W = \sigma^2/S^2$, then $S^2 dw = d\sigma^2$, $0 < W < +\infty$ and

$$\begin{aligned}
&\int_0^{+\infty} \left(\frac{\sigma^2}{S^2}\right)^{1-a} \left(\frac{\sigma^2}{S^2}\right)^{\frac{\nu}{2}-1} \left(1 + \frac{\nu}{n_p}\frac{\sigma^2}{S^2}\right)^{-\frac{\nu_1}{2}} d\sigma^2 \\
&= S^2 \int_0^{+\infty} W^{1-a} W^{\frac{\nu}{2}-1} \left(1 + \frac{\nu}{n_p} W\right)^{-\frac{\nu_1}{2}} dw \\
&= S^2 \frac{\Gamma(\frac{n_p}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{n_p+\nu}{2})} \left(\frac{n_p}{\nu}\right)^{\frac{\nu}{2}} E\left\{W^{-(a-1)}\right\},
\end{aligned}$$

where $E\{.\}$ corresponds to the expectation of the $F(\nu, n_p)$ distribution. Let $W^* = W^{-1}$, then $W^* \sim F(n_p, \nu)$ and

$$E\left\{(W^*)^{a-1}\right\} = E\left\{W^{-(a-1)}\right\} = \left(\frac{\nu}{n_p}\right)^{a-1} \frac{\Gamma(\frac{n_p}{2}+a-1)\Gamma(\frac{\nu}{2}-a+1)}{\Gamma(\frac{n_p}{2})\Gamma(\frac{\nu}{2})},$$

for $a < \dfrac{\nu}{2} + 1$. Thus,

$$\int_0^{+\infty} \left(\frac{\sigma^2}{S^2}\right)^{1-a} \left(\frac{\sigma^2}{S^2}\right)^{\frac{\nu}{2}-1} \left(1 + \frac{\nu}{n_p}\frac{\sigma^2}{S^2}\right)^{-\frac{\nu_1}{2}} d\sigma^2$$

$$= \quad S^2\left(\frac{\nu}{n_p}\right)^{\frac{\nu}{2}-a+1}\frac{\Gamma(\frac{n_p}{2}+a-1)\Gamma(\frac{\nu}{2}-a+1)}{\Gamma(\frac{n_p+\nu}{2})}. \tag{2.4}$$

By using Equation (2.4), we have that

$$\begin{aligned}\pi(\phi,\nu|\mathbf{y}) \quad \propto \quad & \frac{\nu^{\frac{(\nu+n)}{2}}g(\nu)\left|\mathbf{V}_{\hat{\beta}}\right|^{\frac{1}{2}}}{\nu_1^{\frac{\nu_1+k}{2}}g(\nu_1)\,|\mathbf{R}|^{\frac{1}{2}}}\frac{\nu_1^{\frac{\nu_1}{2}}}{n_p^{\frac{\nu_1}{2}}}\left(\frac{n_p}{\nu}\right)^{\frac{\nu}{2}-a+1} \\ & \frac{\Gamma(\frac{n_p}{2}+a-1)\Gamma(\frac{\nu}{2}-a+1)}{\Gamma(\frac{n_p+\nu}{2})}(S^2)^{-(\frac{n_p}{2}+a-1)}\pi(\phi,\nu),\end{aligned}$$

for $a < \nu/2+1$. Using the fact that $g(\nu)/g(\nu_1) = \nu_1^{\frac{p}{2}}\Gamma(\dfrac{\nu_1}{2})/(\nu^{\frac{n}{2}}\Gamma(\dfrac{\nu}{2}))$, and after some algebraic manipulations, we have finally that

$$\begin{aligned}\pi(\phi,\nu|\mathbf{y}) \quad & \propto \quad \nu^{-(1-a)}\frac{\Gamma(\frac{\nu}{2}-a+1)}{\Gamma(\frac{\nu}{2})}\,|\mathbf{R}|^{-\frac{1}{2}}\left|V_{\hat{\beta}}\right|^{\frac{1}{2}}(n_pS^2)^{-(\frac{n_p}{2}+a-1)}\pi(\phi,\nu) \\ & \propto \quad A(\nu)\,|\mathbf{R}|^{-\frac{1}{2}}\left|V_{\hat{\beta}}\right|^{\frac{1}{2}}\left\{(n_p)S^2\right\}^{-(\frac{n_p}{2}+a-1)}\pi(\phi,\nu) \\ & \propto \quad \mathbf{A}(\nu)L^*(\phi,\mathbf{y})\pi(\nu,\phi)\end{aligned}$$

where $A(\nu) = \nu^{-(1-a)}\Gamma\left(\dfrac{\nu}{2}-a+1\right)/\Gamma\left(\dfrac{\nu}{2}\right)$. So, the result follows. □

2.2.2 Proposal of $\pi(\phi,\nu)$

Rewrite $\theta = (\boldsymbol{\beta},\theta^*)$, with $\theta^* = (\sigma^2,\phi,\nu)$ and $L(\theta|\mathbf{y})$ the sampling distribution defined in (1.2). For the *reference* prior, θ^* is the parameter of interest and we assume that $\boldsymbol{\beta}$ is a nuisance parameter. Now, factorizing the prior distribution $\pi(\boldsymbol{\beta},\theta^*) = \pi(\boldsymbol{\beta}|\theta^*)\pi(\theta^*)$ and choosing $\pi(\boldsymbol{\beta}|\theta^*) = 1$ as this is the reference prior in Equation (1.2), we have that

$$L_1(\theta^*) = \int_{\mathbb{R}^p} L(\theta|\mathbf{y})\pi(\boldsymbol{\beta}|\theta^*)d\boldsymbol{\beta} \propto \frac{\Gamma(\frac{\nu_1}{2})}{\Gamma(\frac{\nu}{2})}\frac{(\sigma^2)^{-\frac{n-p}{2}}}{\{\nu+(n-p)S^*\}^{\frac{\nu_1}{2}}}|V_{\hat{\beta}}|^{\frac{1}{2}}|\mathbf{R}|^{-\frac{1}{2}}\nu^{\frac{\nu}{2}}, \tag{2.5}$$

where $S^* = {}^{S^2}\!/_{\sigma^2}$ and $\nu_1 = (n-p)+\nu$. It is possible to show that this expression converge to the normal case when $\nu \to +\infty$. Using the prior reference method (Berger & Bernardo J., 1991), it is necessary to calculate $E\left\{\mathbf{Z}^\top\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\phi}\boldsymbol{\Sigma}^{-1}\mathbf{Z}|S^*\right\}$ and $E\left\{\left(\mathbf{Z}^\top\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\phi}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\right)^2|S^*\right\}$, where $\mathbf{Z} = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})$. Unfortunately, these conditional expectations have no analytical form for the Student-t case. One possible solution is to numerically compute these expressions by using Monte Carlo approximation which will demand a high computational cost, making inference infeasible. For this reason, we suggest the use of the marginal expectations $E\left\{\mathbf{Z}^\top\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\phi}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\right\}$ and $E\left\{\left(\mathbf{Z}^\top\boldsymbol{\Sigma}^{-1}\dfrac{\partial\boldsymbol{\Sigma}}{\partial\phi}\boldsymbol{\Sigma}^{-1}\mathbf{Z}\right)^2\right\}$ in our prior proposal. This suggestion may result in a improper prior (Theorem 2.1), but lead to a proper posterior distribution (Theorem 2.2).

Theorem 2.1. *Under the T-SR model defined by (1.1) and (1.2), for $\phi > 0$ and $\nu > 4$, the prior distribution obtained through the reference prior method is of the form (2.1), with $a = 1$ and*

$$\pi^R(\phi,\nu) \propto \left(BCD + 16\left(B_{11}C_{11}B_{12} - BC_{11}^2\right) - 8B_{12}^2C - \frac{1}{2}B_{11}^2D\right)^{\frac{1}{2}}, \tag{2.6}$$

where

$$\begin{aligned} B &= \frac{\nu(n-p)}{\tau(\nu)},\ C = \left(\frac{2(n-p)}{\tau(\nu)\nu}+1\right)A - \frac{\nu+2}{\nu-2}tr^2\left[\Phi\right], \\ D &= -\left(\frac{2(n-p)}{\nu}\frac{\tau(\nu)+2}{\tau(\nu)(\tau(\nu)-2)} + \delta_1(\nu)\right),\ B_{11} = -\frac{2\nu(n-p)}{(\nu-2)\tau(\nu)}tr\left[\Phi\right], \\ B_{12} &= -\frac{n-p}{(\tau(\nu)-2)\tau(\nu)} \text{ and } C_{11} = \frac{n-p}{(\nu-2)(\tau(\nu)-2)\tau(\nu)}tr\left[\Phi\right], \end{aligned}$$

with $\tau(\nu) = n - p + \nu + 2$, $\Phi = \dfrac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}$, $\mathbf{P} = I - \mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}$ $A = \dfrac{\nu^2}{(\nu-2)(\nu-4)}\left(2tr\left[\Phi^2\right] + tr^2\left[\Phi\right]\right)$, *and* $\delta_1(\nu) = \Psi_1(\nu+n-p/2) - \Psi_1(\nu/2)$, *where* $\Psi_1(.)$ *denotes the trigamma function.*

More details about the proof along with the proof that $\pi^R(\phi,\nu) > 0$ are available in Appendices A3 and A4. The following Lemma provides conditions to prove the results of Theorem 2.2.

Lemma 2.1. *Consider $h = \nu - 4$ and $\mathbf{1} = (1,\ldots,1)^\top$ being an $n \times 1$ vector of ones. Under conditions presented in the Appendix A1.1, we have that:*

- *If* $\mathbf{1}$ *is not a column of* $\mathbf{X}$. *Then,*
  - *(a) as* $(\phi,h) \to (+\infty,+\infty)$,
  $$\pi^R(\phi,\nu) = O\left(h^{-\frac{3}{2}}\frac{d}{d\phi}\log\psi(\phi)\right);$$
  - *(b) as* $(\phi,h) \to (+\infty,0)$,
  $$\pi^R(\phi,\nu) = O\left(h^{-\frac{1}{2}}\frac{d}{d\phi}\log\psi(\phi)\right).$$
- *If* $\mathbf{1}$ *is a column of* $\mathbf{X}$. *Then,*
  - *(a) as* $(\phi,h) \to (+\infty,+\infty)$,
  $$\pi^R(\phi,\nu) = O\left(h^{-\frac{3}{2}}\frac{\omega(\phi)}{\psi(\phi)}\frac{d}{d\phi}\log\left[\frac{\omega(\phi)}{\psi(\phi)}\right]\right);$$
  - *(b) as* $(\phi,h) \to (+\infty,0)$,
  $$\pi^R(\phi,\nu) = O\left(h^{-\frac{1}{2}}\frac{\omega(\phi)}{\psi(\phi)}\frac{d}{d\phi}\log\left[\frac{\omega(\phi)}{\psi(\phi)}\right]\right),$$

where $\psi(\phi) = \nu(\phi)$ and $\omega(\phi)$ are as defined in Berger et al. (2001) for each correlation matrix discussed in Section 1.1.1.

Proof. Consider $h = \nu - 4$. Under the conditions of Section A1.1, we have that,

1. If $\mathbf{1}$ is not a column of $\mathbf{X}$. Then,

$$tr\left[\Phi\right] = O\left(\frac{d}{d\phi}\log\psi(\phi)\right);$$

2. If $\mathbf{1}$ is not a column of $\mathbf{X}$. Then,

$$tr\left[\Phi\right] = O\left(\frac{\omega(\phi)}{\psi(\phi)}\frac{d}{d\phi}\log\left[\frac{\omega(\phi)}{\psi(\phi)}\right]\right).$$

Next, we will proof the case when $\mathbf{1}$ is not a column of $\mathbf{X}$, the case when $\mathbf{1}$ is a column of $\mathbf{X}$ is analogous and will be omitted. Let $n_p = n - p$ and using the Stirling approximation for the trigamma function ($\Psi_1(x) = x^{-1} + (2x^2)^{-1}$), we have that

$$D \simeq \frac{1}{4}\left\{\frac{-2n_p^3\nu + 2n_p^3 - 2n_p^2\nu^2 - 4n_p^2\nu + 8n_p^2 + 8n_p\nu + 8n_p}{\nu^2(n_p+\nu^2)(n_p+\nu+2)}\right\}, \quad \text{as} \quad \nu \to \infty,$$

where $\simeq$ denotes assymptotic equality. Note that, the higher order for the numerator (in terms of ν) is 2, while the denominator is a polynomial of order 5, so we have that $D = O(\nu^{-3})$ and,

$$BCD = O(1)\left(O(1)A - O(1)tr^2\left[\Phi\right]\right)O(\nu^{-3}).$$

Now, $A = O(1)\left(2tr\left[\Phi^2\right] + tr^2\left[\Phi\right]\right)$. Then, as $\phi \to +\infty$, we have that

$$BCD = O(\nu^{-3})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right).$$

Analogously,

$$\begin{aligned} B_{11}C_{11}B_{12} &= \frac{2\nu n_p^2}{(\nu+n_p+2)^3(\nu-2)^2(\nu+n_p)^2}tr^2\left[\Phi\right] \\ &= O(\nu^{-6})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right), \end{aligned}$$

and hence

$$\begin{aligned} |I_1(\theta^*)| &< BCD + B_{11}C_{11}B_{12} \\ &= O(\nu^{-3})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right) + O(\nu^{-6})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right) \\ &= O(\nu^{-3})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right) \end{aligned}$$

Table 2 – Conditions to guarantee the propriety of the posterior distribution using the proposal prior.

Correlation Family	1 is a column of X	1 is not a column of X
Spherical	$-1 < a < \frac{\nu}{2} + 1$	$\frac{1}{2} < a < \frac{\nu}{2} + 1$
Power exponential	$0 < a < \frac{\nu}{2} + 1$	$\frac{1}{2} < a < \frac{\nu}{2} + 1$
Cauchy	$0 < a < \frac{\nu}{2} + 1$	$\frac{1}{2} < a < \frac{\nu}{2} + 1$
Matern ($\kappa < 1$)	$2 - \kappa^{-1} < a < \frac{\nu}{2} + 1$	$\frac{1}{2} < a < \frac{\nu}{2} + 1$

$$\equiv \quad O(h^{-3})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right).$$

Finally, when $h \to 0$ (or $\nu \to 4$) we have,

$$BCD + B_{11}C_{11}B_{12} = O(1)\left(A - O(1)tr^2\left[\Phi\right]\right)O(1) + O(1)tr^2(\Phi)$$

, where, $A = O\left(1/h\right)\left(2tr\left[\Phi^2\right] + tr^2\left[\Phi\right]\right)$, so, as we have $1/h$ grows faster than a constant function when $h \to 0$, then,

$$BCD + B_{11}C_{11}B_{12} = O\left(\frac{1}{h}\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right).$$

and the result follows. □

The next Theorem 2.2 shows the conditions under which the prior in (2.1) generates a proper posterior using $\pi(\phi,\nu) = \pi^R(\phi,\nu)$ as described in Theorem 2.1, the proposed prior is a particular case considering $a = 1$.

Theorem 2.2. *Using the prior defined in* (2.1), $\pi(\phi,\nu) = \pi^R(\phi,\nu)$ *and under the T-SR model* (1.1), *the posterior distribution of* $(\boldsymbol{\beta},\sigma,\phi,\nu)$ *is proper if the conditions in Table 2 are satisfied for the hyperparameter a. This holds for any of the families of correlation functions considered in Section 1.1.1.*

Proof. Let,

$$L^*(\phi,\mathbf{y}) = |\mathbf{R}|^{-\frac{1}{2}}\left|V_{\hat{\beta}}\right|^{\frac{1}{2}}\left((n-p)S^2\right)^{-\left(\frac{n-p}{2}+a-1\right)} \tag{2.7}$$

and $\pi(\phi,\nu) = \pi^R(\phi,\nu)$. To guarantee the propriety of the posterior distribution condition (2.3) must be satisfied. To show this, we have to proof integrability of (2.2) at the integration limits for both ϕ and ν. When $\phi \to 0$, $\boldsymbol{\Sigma} \to I$, so, we have that,

$$A(h)L^*(\phi,\mathbf{y}),$$

is bounded when $h \to 0$ and $h \to +\infty$. This is true because $A(h)$ has a constant behavior when $h \to 0$, while, when $h \to +\infty$ we have that $A(h) \to 2^{1-a}$ (Appendix A2). With respect to $\pi^R(\phi, \nu)$, following the first paragraph of the proof of Theorem 4 in Berger *et al.* (2001), we can conclude that we also have integrability of $\left(tr^2\left[\partial\mathbf{\Sigma}/\partial\phi\right]\right)^{1/2}$ when $\phi \to 0$. Using this fact and Lemma 2.1, we have that for a small M,

$$\int_0^M \pi^R(\phi, h)d\phi = O\left(h^{-\frac{3}{2}}\right), \qquad \text{when } h \to +\infty,$$
$$\int_0^M \pi^R(\phi, h)d\phi = O\left(h^{-\frac{1}{2}}\right), \qquad \text{when } h \to 0,$$

thus, $\pi^R(\phi, \nu)$ is integrable when $(\phi, h) \to (0, +\infty) \equiv (\phi, \nu) \to (0, +\infty)$ and $(\phi, h) \to (0, 0) \equiv (\phi, \nu) \to (0, 4)$, since we have that $h^{-\frac{3}{2}}$ and $h^{-\frac{1}{2}}$ are integrable functions at $+\infty$ and 0, respectively. Then we can conclude that (2.2) is integrable when $(\phi, \nu) \to (0, 4)$ and $(\phi, \nu) \to (0, +\infty)$. Now, when $\phi \to +\infty$, $\nu \to +\infty$ or $\nu \to 4$, by Lemma 2.1 and Lemma 1 from Berger *et al.* (2001), if $\mathbf{1}$ is a column of $\mathbf{X}$ we have that,

$$\int_0^{+\infty} A(h)L^*(\phi, \mathbf{y})\pi^R(\phi, h)dh \leqslant C\psi(\phi)^{a-1}\left|\frac{\omega(\phi)}{\psi(\phi)}\frac{d}{d\phi}\log\frac{\omega(\phi)}{\psi(\phi)}\right|, \tag{2.8}$$

while if $\mathbf{1}$ is not a column of $\mathbf{X}$,

$$\int_0^{+\infty} A(h)L^*(\phi, \mathbf{y})\pi^R(\phi, h)dh \leqslant C\psi(\phi)^{a-\frac{1}{2}}|\log\psi(\phi)|, \tag{2.9}$$

with C a limiting constant. The previous reasoning is true because $h^{-\frac{3}{2}}$ and $h^{-\frac{1}{2}}$ are integrable at $+\infty$ and 0, respectively. Therefore, expressions (2.8) and (2.9) coincides with the boundaries of the normal case which has been proven to be integrable when $\phi \to +\infty$ (Berger *et al.*, 2001). With all these facts, it follows that (2.2) is integrable when $(\phi, \nu) \to (+\infty, 4)$ and $(\phi, \nu) \to (+\infty, +\infty)$. □

Muré (2020) has shown that under the normal distribution, the Matérn family with $\kappa \geqslant 1$ does not fulfill all the conditions to generate proper posteriors when $\pi^R(\phi)$ is used. As the correlation matrix structures considered for the T-SR are the same as the normal one, the results from Muré (2020) are also applicable under the T-SR. Therefore, for the T-SR, the Matérn family with $\kappa \geqslant 1$ does not generate proper posteriors and was not included in Table 2.

The following section states the Jeffreys priors which are built using the information matrix. Here, we establish both the Jeffreys rule and the Jeffreys independence densities giving conditions under which they generate proper posteriors distributions. These priors will be used in simulation studies and a real data application as comparison points for the proposal procedure.

2.3 The Jeffreys Prior

Let $\ell(\theta|\mathbf{y}) = \log(L(\boldsymbol{\theta}|\mathbf{y}))$ be the log-likelihood (1.2) for the T-SR model, with $\theta = (\boldsymbol{\beta}, \sigma^2, \phi, \nu)$. Using the information matrix $I(\theta)$ from Lange *et al.* (1989), the Jeffreys-rule prior for model (1.1) is given by $\pi(\theta) \propto \sqrt{|I(\theta)|}$, with

$$I(\theta) = E_\theta \left\{ \left(\frac{\partial \ell(\theta|\mathbf{y})}{\partial \theta} \right) \left(\frac{\partial \ell(\theta|\mathbf{y})}{\partial \theta} \right)^\top \right\},$$

where $E_\theta \{.\}$ is the conditional expectation of $\mathbf{Y}$ given a value of θ. The independence Jeffreys prior can be derived from $I(\theta)$ assuming that $\boldsymbol{\beta}$ and $\theta^* = (\sigma^2, \phi, \nu)$ are independent a priori.

Theorem 2.3. *Let* $\tau_1(\nu) = n + \nu + 2, \Phi_1 = \dfrac{\partial \mathbf{R}}{\partial \phi} \mathbf{R}^{-1}$ *and* $\Psi_1(.)$ *the trigamma function. Under the T-SR model* (1.1)*.*

1. *The independence Jeffreys prior denoted by* $\pi^{JI}(\boldsymbol{\beta}, \sigma, \phi, \nu)$ *is of the form (4.22), with* $a = 1$, $\phi > 0$, $\nu > 4$ *and*

$$\pi^{JI}(\phi, \nu) \propto \left(B^* C^* D^* + 8 B^*_{11} B^*_{12} C^*_{11} - 4 B^{*2}_{12} C - 4 C^{*2}_{11} B - B^{*2}_{11} D^* \right)^{\frac{1}{2}}, \qquad (2.10)$$

where

$$\begin{aligned} B^* &= \frac{n(\nu+2)}{\tau_1(\nu)}, \; C^* = \frac{1}{\tau_1(\nu)} \left\{ (\tau_1(\nu) - 2) tr \left[\Phi_1^2 \right] - tr^2 \left[\Phi_1 \right] \right\}, \\ D^* &= -\frac{1}{2} \left\{ \frac{1}{2} \Psi_1 \left(\frac{\tau_1(\nu) - 2}{2} \right) - \frac{1}{2} \Psi_1 \left(\frac{\nu}{2} \right) + \frac{n(\tau_1(\nu) + 2)}{\nu(\tau_1(\nu) - 2)(\tau_1(\nu))} \right\}, \\ B^*_{11} &= \frac{\tau_1(\nu) - 3}{\tau_1(\nu)} tr \left[\Phi_1 \right], B^*_{12} = \frac{n}{\tau_1(\nu)(\tau_1(\nu) + 2)} \; \textit{and} \\ C^*_{11} &= \frac{1}{\tau_1(\nu)(\nu + 2)} tr \left[\Phi_1 \right], \end{aligned}$$

2. *The Jeffreys-rule prior denoted by* $\pi^J(\boldsymbol{\beta}, \sigma, \phi, \nu)$, *is of the form* (4.22) *with* $a = {}^{p}/_{2} + 1$, $\nu > \max\{p, 4\}$, *and*

$$\pi^J(\phi, \nu) \propto \left(\frac{\tau_1(\nu) - 2}{\tau_1(\nu)} \right)^{\frac{p}{2}} \left| V_{\hat{\beta}} \right|^{-\frac{1}{2}} \pi^{JI}(\phi, \nu), \qquad (2.11)$$

where $\tau_1(\nu)$ *and* $\pi^{JI}(\phi, \nu)$ *are given in item (1).*

Proof. Let $\ell(\theta)$ be the log-likelihood of $\theta = (\boldsymbol{\beta}, \theta^*)$ for the T-SR model, with $\theta^* = (\sigma^2, \phi, \nu)$. As in Lange *et al.* (1989), let us define $\mu = \mathbf{X}\boldsymbol{\beta}$ and $\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}$. Then, the Fisher information matrix can be written as:

$$I(\theta) = E \left\{ \left(\frac{\partial \ell(\theta)}{\partial \theta} \right) \left(\frac{\partial \ell(\theta)}{\partial \theta} \right)^\top \right\} = \begin{bmatrix} I_{\beta\beta^\top} & \mathbf{0} \\ \mathbf{0}^\top & I_{\theta*} \end{bmatrix} = \begin{bmatrix} I_{\beta\beta^\top} & \underline{\mathbf{0}} & \underline{\mathbf{0}} & \underline{\mathbf{0}} \\ \underline{\mathbf{0}}^\top & I_{(\sigma^2)^2} & I_{\sigma^2\phi} & I_{\sigma^2\nu} \\ \underline{\mathbf{0}}^\top & I_{\sigma^2\phi} & I_{(\phi)^2} & I_{\phi\nu} \\ \underline{\mathbf{0}}^\top & I_{\sigma^2\nu} & I_{\phi\nu} & I_{\nu^2} \end{bmatrix},$$

where $\mathbf{0}$ is a $p \times 3$ matrix of zeros, $\underline{\mathbf{0}}$ is a 3-dimensional vector of zeros and:

$$
\begin{aligned}
I_{\beta\beta^\top} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \ell(\theta)}{\partial \boldsymbol{\beta}}\right)^\top\right\} = \frac{1}{\sigma^2}\frac{\tau_1(\nu)-2}{\tau_1(\nu)}\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X},\\
I_{(\sigma^2)^2} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \sigma^2}\right)^2\right\} = \frac{1}{2\sigma^4}\frac{n(\nu+2)}{\tau_1(\nu)} = \frac{1}{2\sigma^4}B^*,\\
I_{\phi^2} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \phi}\right)^2\right\} = \frac{\tau_1(\nu)-2}{2\tau_1(\nu)}tr\left[\Phi_1^2\right] - \frac{1}{2\tau_1(\nu)}tr^2\left[\Phi_1\right] = \frac{1}{2}C^*,\\
I_{\nu^2} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \nu}\right)^2\right\}\\
&= -\frac{1}{2}\left\{\frac{1}{2}\Psi_1\left(\frac{\tau_1(\nu)-2}{2}\right) - \frac{1}{2}\Psi_1\left(\frac{\nu}{2}\right) + \frac{n(\nu+n+4)}{\nu(\nu+n)(\nu+n+2)}\right\} = \frac{1}{2}D^*,\\
I_{\sigma^2\phi} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \sigma^2}\right)\left(\frac{\partial \ell(\theta)}{\partial \phi}\right)\right\} = \frac{1}{2\sigma^2}\frac{\tau_1(\nu)-3}{\tau_1(\nu)}tr\left[\Phi_1\right] = \frac{1}{2\sigma^2}B^*_{11},\\
I_{\sigma^2\nu} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \sigma^2}\right)\left(\frac{\partial \ell(\theta)}{\partial \nu}\right)\right\} = -\frac{n}{\sigma^2\tau_1(\nu)(\nu+2)} = \frac{1}{\sigma^2}B^*_{12},\\
I_{\phi\nu} &= E\left\{\left(\frac{\partial \ell(\theta)}{\partial \phi}\right)\left(\frac{\partial \ell(\theta)}{\partial \nu}\right)\right\} = -\frac{1}{\tau_1(\nu)(\nu+2)}tr\left[\Phi_1\right] = \frac{1}{2}C^*_{11},
\end{aligned}
$$

with $\tau_1(\nu) = n+\nu+2, \Phi_1 = \dfrac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}$ and $\Psi_1(.)$ denoting the trigamma function. For the independence Jeffreys prior, we have that $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \nu) \propto \pi(\boldsymbol{\beta}) \times \pi(\theta^*)$, where it is assumed that $\pi(\boldsymbol{\beta}) \propto 1$ and $\pi(\theta^*) \propto \sqrt{|I_{\theta*}|}$. After some algebra, we obtain the prior defined in equation (4.22) of section 2.2.1 in the main version of the manuscript, with $a = 1$ and $\pi(\phi, \nu)$ as defined in Equation (2.10) in section 2.3.

For the Jeffreys-rule case, we take advantage of the block diagonal structure of $I(\theta)$ obtaining $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \nu) \propto \sqrt{|I_{\beta\beta^\top}| \times |I_{\theta*}|}$. After some algebraic manipulation, we obtain the prior defined in equation (4.22) of section 2.2.1, with $a = p/2 + 1$ and $\pi(\phi, \nu)$ as defined in Equation (2.10) in section 2.3. □

As we can see, the Jeffreys prior is the resulting product between the independence Jeffreys prior with another function that depends on ϕ and ν. In the following Lemma, besides analyzing $\pi^{JI}(\phi, \nu)$ at $\nu \to +\infty$, we also consider the case $\nu \to max(4, p)$, this will be useful to show propriety of the Jeffreys prior ($a = p/2 + 1$) since there is a discontinuity point at $\nu = p$, whereby, the integrability property could not be assured.

Lemma 2.2. *Let $p^* = \max(p, 4)$, with $u = \nu - p^*$. Under conditions introduced in the Appendix A1.1, we have that*

1. *as $u \to 0$ and $\phi \to +\infty$,*

$$\pi^{JI}(\phi, \nu) = O\left(u^{-\frac{1}{2}}\frac{d}{d\phi}\log\psi(\phi)\right); \tag{2.12}$$

2. as $u \to +\infty$ and $\phi \to +\infty$,

$$\pi^{JI}(\nu,\phi) = O\left(u^{-2}\frac{d}{d\phi}\log\psi(\phi)\right). \tag{2.13}$$

Proof. As we know, the Jeffreys prior is the resulting product between the Jeffreys independent prior with another functions that depends on ϕ and ν. In this section, we analize the behaviour of $\pi^{JI}(\phi,\nu)$ at $u = \nu - p^*$ since for the Jefreys prior, there appears to be a discontinuity at $\nu = p$, note first that,

$$\pi^{JI}(\nu,u) < (B^*C^*D^* + 16B^*_{11}B^*_{12}C^*_{11})^{\frac{1}{2}}.$$

For fixed n, as $\nu \to p^*$, we have that

$$\begin{aligned}
\frac{(n(u+p^*+2)(u+p^*+n))}{\tau^2(u)} &= O(u), & \frac{n(u+p^*+2)}{\tau^2(u)} &= O(u),\\
\Psi_1(u) &= O(u^{-2}), & \frac{\tau(u+p^*)-3}{\tau(u)} &= O(u),\\
\frac{n}{(\tau(u)(\tau(u)+2))} &= O(u^{-1}), & \frac{1}{\tau(u)(u+2)} &= O(u^{-1}),
\end{aligned}$$

Then, as $\phi \to +\infty$,

$$(B^*C^*D^* + B^*_{11}B^*_{12}C^*_{11})^{1/2} = O\left(u^{-\frac{1}{2}}\frac{d}{d\phi}\log\psi(\phi)\right),$$

since $\Phi_1 = O\left(\frac{d}{d\phi}\log\psi(\phi)\right)$. Thus, by the transitive property of real numbers

$$\pi^{JI}(\phi,\nu) = O\left(u^{-\frac{1}{2}}\frac{d}{d\phi}\log\psi(\phi)\right). \tag{2.14}$$

On the other hand, as $\nu \to +\infty$, by using the Stirling approximation, we have that

$$D^* \approx \frac{-n^2\nu - 4n\nu - n^3 - 2n^2}{\nu^2(\nu+n)^2(\nu+n+2)} = O(\nu^{-4}).$$

Also, $B^* \to 1$, $C^* = O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right)$ and

$$B^*_{11}B^*_{12}C^*_{11} = O(\nu^{-4})O\left(\left(\frac{d}{d\phi}\log\psi(\phi)\right)^2\right),$$

hence

$$\pi^{JI}(\nu,\phi) = O\left(\nu^{-2}\frac{d}{d\phi}\log\psi(\phi)\right) \equiv O\left(u^{-2}\frac{d}{d\phi}\log\psi(\phi)\right). \tag{2.15}$$

□

Moreover, in the following theorems, we present the conditions under which the independence Jeffreys and the Jeffreys-rule priors have a proper posterior distribution.

Theorem 2.4. *Consider the T-SR model* (1.1) *with the families of correlation functions listed in Section 1.1.1. Then, when* $\mathbf{1}$ *is not a column of* $\mathbf{X}$*, the independence Jeffreys prior* $\pi^{JI}(\boldsymbol{\beta}, \sigma, \phi, \nu)$*, yields a proper posterior distribution, while when* $\mathbf{1}$ *is a column of* $\mathbf{X}$*, the independence Jeffreys prior yields an improper posterior distribution.*

Proof. As for the reference case, we have to study the behavior of (2.2) at the integration limits. We already explained in proof of Theorem 2.2 that $L^*(\phi, \mathbf{y})$ is bounded when $\phi \to 0$, on the other hand, since $\left(tr^2\left[\frac{\partial \boldsymbol{\Sigma}}{\partial \phi}\right]\right)^{1/2}$ is integrable at zero, using Lemma 2.2, for a small enough M, we have that,

$$\int_0^M \pi^{JI}(\phi, u)d\phi = O\left(u^{-\frac{1}{2}}\right), \text{ when } u \to 0,$$
$$\int_0^M \pi^{JI}(\phi, u)d\phi = O\left(u^{-2}\right), \text{ when } u \to \infty,$$

thus, $\pi^R(\phi, \nu)$ is integrable when $(\phi, u) \to (0, 0) \equiv (\phi, \nu) \to (0, p^*)$ and $(\phi, u) \to (0, +\infty) \equiv (\phi, \nu) \to (0, +\infty)$, since we have $u^{-1/2}$ and u^{-2} are integrable functions at 0 and $+\infty$, respectively. Then Equation (2.2) is integrable when $(\phi, \nu) \to (0, p^*)$ and $(\phi, \nu) \to (0, +\infty)$. Now, when $\phi \to +\infty$, $u \to 0$ or $u \to +\infty$, by Lemma 2.1 , Lemma 1 and 3 from Berger *et al.* (2001), if $\mathbf{1}$ is a column of $\mathbf{X}$ we have that,

$$\begin{aligned} \int_0^{+\infty} A(u)L^*(\phi, \mathbf{y})\pi^{JI}(\phi, \nu)du &\leqslant \psi(\phi)^{1-1}C|\log \psi(\phi)| \\ &\leqslant C|\log \psi(\phi)|, \end{aligned} \tag{2.16}$$

while if $\mathbf{1}$ is not a column of $\mathbf{X}$,

$$\begin{aligned} \int_0^{+\infty} A(u)L^*(\phi, \mathbf{y})\pi^{R}(\phi, u)du &\leqslant C\psi(\phi)^{1-\frac{1}{2}}|\log \psi(\phi)| \\ &\leqslant C\psi(\phi)^{\frac{1}{2}}|\log \psi(\phi)|, \end{aligned} \tag{2.17}$$

with C a limiting constant. Because $u^{-1/2}$ and u^{-2} are integrable functions at 0 and $+\infty$ respectively, we conclude that the previous reasoning is true. Like in the proof of Theorem 2.2, expressions (2.16) and (2.17) coincide with the boundaries of the normal case which were already explored by Berger *et al.* (2001) for all the correlation functions considered. Therefore, the result follows. □

Theorem 2.5. *For the T-SR model* (1.1) *and any family of correlation functions considered in Section 1.1.1, the Jeffreys-rule prior* $\pi^J(\boldsymbol{\beta}, \sigma, \phi, \nu)$ *yields a proper posterior distribution.*

Proof. First, note that for the Jeffreys prior we have $a = \dfrac{p}{2} + 1$ then,

$$A(u)L(\phi, \mathbf{y})\pi^J(\phi, \nu) = A(u)\left(\frac{\tau(u)-2}{\tau(u)}\right)^{\frac{p}{2}} |\mathbf{R}|^{-\frac{1}{2}} \left((n-p)S^2\right)^{-\frac{n}{2}} \pi^{JI}(\phi, u),$$

considering the change of variable $u = \nu - p^*$, For the Jeffreys-rule prior, we have to prove that

$$\int_0^{+\infty}\int_0^{+\infty} A(u)\left(\frac{\tau(u)-2}{\tau(u)}\right)^{\frac{p}{2}} |\mathbf{R}|^{-\frac{1}{2}} \left((n-p)S^2\right)^{-\frac{n}{2}} \pi^{JI}(\phi, u) du d\phi < +\infty. \qquad (2.18)$$

Now, we proceed to analyze each term of the integral (2.18). As $\phi \to +\infty$, we have that

$$\Phi_1 = O\left(\frac{d}{d\phi}\log\psi(\phi)\right), \quad |\mathbf{R}|^{-\frac{1}{2}} = O(\psi(\phi)^{n-1}) \quad \text{and} \quad S^2 = O(\psi(\phi)^{-1}),$$

also when $u \to +\infty$ we have, $\dfrac{\tau(u)-2}{\tau(u)} = O(1)$ and $A(u) = O(1)$ while when $u \to 0$, $\dfrac{\tau(u)-2}{\tau(u)} = O(u)$ and $A(u) = O(u^{-1})$, then, by Lemma 2.2

$$A(u)L(\phi, \mathbf{y})\pi^J(\phi, u) = O\left(u^{\frac{p-3}{2}}\frac{d}{d\phi}\log\phi\right), \text{ when } u \to 0, \qquad (2.19)$$

$$A(u)L(\phi, \mathbf{y})\pi^J(\phi, u) = O\left(u^{-2}\frac{d}{d\phi}\log\phi\right), \text{ when } u \to \infty. \qquad (2.20)$$

Hence, when $(\phi, u) \to (+\infty, 0) \equiv (\phi, \nu) \to (+\infty, p^*)$ or $(\phi, u) \to (+\infty, \infty) \equiv (\phi, \nu) \to (+\infty, \infty)$ we have that

$$\int_0^{+\infty} A(u)L(\phi, \mathbf{y})\pi^J(\phi, u) du = O\left(\frac{d}{d\phi}\log\phi\right). \qquad (2.21)$$

Since $u^{\frac{p-3}{2}}$ and u^{-2} are integrable fuctions at zero and $+\infty$, respectively, then (2.18) is integrable at $(\phi, \nu) \to (+\infty, p^*)$ and $(\phi, \nu) \to (+\infty, +\infty)$. Similarly to the previous priors, we have that $\left(tr^2\left[\dfrac{\partial\mathbf{\Sigma}}{\partial\phi}\right]\right)^{1/2}$ is integrable at zero and because of the integrability of $u^{\frac{p-3}{2}}$ and u^{-2} at the points of interest, we conclude that (2.18) is integrable at $(\phi, \nu) \to (0, p^*)$ and $(\phi, \nu) \to (0, +\infty)$. □

Having established the priors we use through this work jointly with conditions under which they generate proper posteriors, we proceed to give details about the sampling procedure we used to make the Bayesian inferences about the considered model.

2.4 Bayesian Inference

In this section, we present our Gibbs sampler algorithm with a metropolis step. In addition, the model selection metrics are also introduced.

2.4.1 Gibbs sampling

Using the properties given in Kotz & Nadarajah (2004) for the Student-t distribution, we can generalize the results of Zellner (1976) considering a covariance matrix $\mathbf{\Sigma} = \sigma^2\mathbf{R}$. Given the model in (1.1) and conditioning on the values of (ϕ, ν), the posterior density of $(\boldsymbol{\beta}, \sigma^2)$ is given by

$$p(\boldsymbol{\beta}, \sigma^2|\phi, \nu, \mathbf{Z}) \propto \frac{(\sigma^2)^{(\frac{\nu}{2}-1)}}{(\tilde{\sigma}^2)^{\frac{\nu+n}{2}}}\left[\nu_1 + \frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^\top V_{\hat{\boldsymbol{\beta}}}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{\tilde{\sigma}^2}\right]^{-\left(\frac{\nu+n}{2}\right)}, \tag{2.22}$$

where $\tilde{\sigma}^2 = \sigma^2\nu + n_p S^2/\nu_1$, $\nu_1 = n_p + \nu$, $n_p = n - p$, $S(\phi)^2 = S^2 = (\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top\mathbf{R}_\phi^{-1}(\mathbf{Z} - \mathbf{X}\hat{\boldsymbol{\beta}}))/n_p$, $\hat{\boldsymbol{\beta}}(\phi) = \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top\mathbf{R}_\phi^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{R}_\phi^{-1}\mathbf{y}$ and $V_{\hat{\boldsymbol{\beta}}} = (X^\top\mathbf{R}_\phi^{-1}X)^{-1}$. Using (2.22), it can be shown that $\boldsymbol{\beta}|\phi, \mathbf{y} \sim t_p(\hat{\boldsymbol{\beta}}, S^2V_{\hat{\boldsymbol{\beta}}}, n_p)$ and $\frac{\sigma^2}{S^2}|\phi, \nu, \mathbf{y} \sim F(\nu, n_p)$. Therefore, we propose Algorithm 4, a Gibbs sampler with a metropolis step, to sample from the conditional distributions.

Algorithm 4 Obtaining a random sample from $\pi(\boldsymbol{\beta}, \sigma^2, \phi, \nu|\mathbf{y})$ using Gibbs Sampling with a metropolis step.

1: Sample $(\phi^{(s)}, \nu^{(s)})$:
- Generate a candidate $(\nu^\star, \phi^\star)$ from an independent proposal $q(\nu, \phi) = q(\nu)q(\phi)$.
- Calculate $\alpha = \dfrac{\pi(\phi^\star, \nu^\star|\mathbf{y})q(\phi^{(s-1)}, \nu^{(s-1)})}{\pi(\phi^{(s-1)}, \nu^{(s-1)}|\mathbf{y})q(\phi^\star, \nu^\star)}$.
- Sample $U \sim \text{Uniform}(0, 1)$.
- Set $(\phi^{(s)}, \nu^{(s)}) = (\phi^\star, \nu^\star)$ if $u < \alpha$ or $(\phi^{(s)}, \nu^{(s)}) = (\phi^{(s-1)}, \nu^{(s-1)})$ otherwise.

2: Sample $W = \sigma^2/S^2(\phi^{(s)}) \sim F(\nu^{(s)}, n_p)$:
- Given $\phi^{(s)}$, compute $S^2(\phi^{(s)})$ and recover $\sigma^{2(s)} = S^2(\phi^{(s)})w$.

3: Sample $\boldsymbol{\beta}^{(s)} \sim t_p\left(\hat{\boldsymbol{\beta}}(\phi^{(s)}), S^2(\phi^{(s)})V_{\hat{\boldsymbol{\beta}}}(\phi^{(s)}), n_p\right)$.

In Algorithm 4, as in Lobo & Fonseca (2020), we use $q(\nu) = \text{TLN}(\nu^{(s-1)}; \sigma, \nu > 4)$, where $\text{TLN}(\mu; \sigma; A)$ denotes the log-normal distribution with location and scale parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, truncated on the interval A; and $q(\phi) = U(a, b)$ with the hyperparameters calibrated to control the metropolis acceptance to be around 30%. Using Algorithm 4, we implement the `OBASpatial` `R` package (Ordoñez *et al.*, 2021) that is available for download at `CRAN` repository. The `R` package is used to run the simulations and the real data analysis.

2.4.2 Model selection

Let us start by setting up the model selection as an hypothesis testing problem (Banerjee *et al.*, 2014; Berger *et al.*, 2001). Thus, replace the usual hypotheses by a candidate parametric model, say m_k, having respective parameter vectors $\boldsymbol{\theta}_{m_k}$. Under the prior density proposal in Equation (4.22) we compute the marginal density for a model m_k as

$$\begin{aligned} m_k(\mathbf{y}) &= \int L(\boldsymbol{\theta}_{m_k})\pi(\boldsymbol{\theta}_{m_k})d\boldsymbol{\theta}_{m_k} \\ &= \int_0^{+\infty}\int_4^{+\infty} A(\nu)\left|\mathbf{R}\right|^{-\frac{1}{2}}\left|V_{\hat{\boldsymbol{\beta}}}\right|^{\frac{1}{2}}\left\{(n-p)S^2\right\}^{-\left(\frac{n-p}{2}+a-1\right)}\pi(\phi,\nu)d\nu d\phi, \end{aligned}$$

where $L(\boldsymbol{\theta}_k)$ is the likelihood (1.2) for the model m_k.To compare q different models m_k, with $k = 1, \ldots, q$, we assign equal prior probabilities to the models. Therefore, the resulting posterior probability for the k-th model is defined by

$$p(m_k \mid \mathbf{y}) = \frac{m_k(\mathbf{y})}{\sum_{j=1}^{q} m_j(\mathbf{y})}.$$

Under this criterion a model with the largest posterior probability is preferable. Another possibility to perform model selection is to choose the model with best prediction power. Suppose that n_0 locations are separated as a validation set. The mean square prediction error (MSPE) of the k-th model is defined by

$$\mathrm{MSPE}_k = \frac{\sum_{i=1}^{n_0}\left\{Y(s_i) - \hat{Y}^{(k)}(s_i)\right\}^2}{n_0},$$

where $\hat{Y}^{(k)}(s_i)$ is the predicted response for observation in location s_i under the k-th model. Thus, the model that minimizes the MSPE is the most suitable under this criterion.

3 Objective Bayesian analysis for geostatistical Student-t processes: applications with simulated and real data

In this chapter we assess the performance of the proposed prior through simulation and a real data application. In Section 3.1, two simulation studies are performed to assess the frequentist properties of the Bayesian estimates under different priors and compare the performance of the Student-t and Gaussian spatial regression. A real data analysis is performed in Section 3.2 to corroborate our findings of Section 3.1 and finally, a brief discussion is presented in Section 3.3. Technical derivations are relegated to Appendix B.

3.1 Simulation Study

We present a two fold simulation study: 1) we study the frequentist properites of the *reference* prior in comparison with the *Jeffreys* and *vague* priors; and 2) we compare the performance of the Student-t versus Gaussian processes under different priors.

3.1.1 Frequentist properties

The study of the frequentist properties of Bayesian inference is of interest to assess and understand the properties of non-informative or default priors (e.g., Stein, 1985; De Oliveira, 2007; Kazianka & Pilz, 2012; Branco *et al.*, 2013; He *et al.*, 2020). Therefore, a simulation study is performed to assess the performance of the proposal method and compare it with both Jeffreys priors: the Jeffreys-rule (*Jef rul*) and the independence Jeffreys (*Jef ind*) and a vague prior (*vague*) of the form (4.22).

We propose two T-SR models with coordinates $\mathbf{s} = (\mathbf{x}_1, \mathbf{x}_2)$ and $\mathbf{R}$ belonging to the Matérn family with $\kappa = 0.5$ (exponential correlation structure), $\sigma^2 = 0.8$ and $\phi = 2$ to study the proposed priors. The first one (**Scenario 1**) is given by,

$$y(s_i) = 10 + \varepsilon_{s_i}, \quad i, \dots, n, \tag{3.1}$$

with $\varepsilon \sim t_n(\mathbf{0}, \sigma^2\mathbf{R}, \nu = 5)$. And, to illustrate the existing intercept and dimension restrictions of the independent Jeffreys and Jeffreys-rule prior respectively, the second T-SR model (**Scenario 2**) is given by

$$y(s_i) = 0 - 2.2x_{1i} + 0.5x_{2i} + 1.7x_{1i}^2 + 2.4x_{2i}^2 + 3.5x_{1i}x_{2i} + \varepsilon_{s_i}. \tag{3.2}$$

Table 3 – Simulation study 1. Coverage probability (and expected log-length) of the variance structure parameters under two simulation scenarios for $n = 100$.

	Prior	σ	ϕ	ν
Scenario 1	*reference*	0.91 (0.921)	0.98 (0.985)	1 (2.55)
	vague	0.351 (-0.102)	0.85 (0.794)	0.86 (2.40)
	Jef rul	0.575 (0.623)	0.777 (0.681)	0.808 (2.55)
Scenario 2	*reference*	0.939 (0.94)	0.93 (1.097)	1 (2.54)
	vague	0.434 (-0.14)	0.909 (0.97)	0.2 (2.43)
	Jef rul	0.376 (-0.44)	0.586 (0.496)	0 (2.41)
	Jef ind	0.495 (0.385)	0.859 (0.904)	0.1 (2.53)

A total of $K = 100$ Monte Carlo simulations were generated for each scenario, the coordinates $\mathbf{s}$ were sampled at $n = 50, 100, 250$ and 500 locations of a regular lattice in $D_{\mathbf{s}} = [0, 10] \times [0, 10]$.

For the *vague* proper prior, we consider $a = 2.1$ and $\pi(\phi, \nu) = \pi(\phi) \times \pi(\nu)$ with $\pi(\phi) = U(0.1, 4.72)$ and $\pi(\nu) = Texp(\lambda; \nu \in \text{A})$, where $\pi(\lambda) = U(0.01, 0.25)$, $\text{A} = [4.1, +\infty)$ and $Texp(\lambda, \nu \in \text{A})$ denote the truncated exponential distribution. The distribution of λ is such that it allows the mean of the prior of ν to vary from 4 to 100. The exponential prior of ν is truncated above 4.1 to guarantee the existence of the Student-t process and the prior of ϕ allow that the distance such that the empirical range, $\text{corr}(s_i, s_j) < 0.05$, varies from 0.30 to 14 (which is the minimum and maximum distance between the locations, respectively).

For the two scenarios, we compute the highest posterior density 95% credible interval for all parameters, based on the four priors. We also compute the coverage probability for each parameter as the number of simulations in such the parameter is inside the credible limits, and the expected log length of each credible interval as the mean of the logarithm of the difference (log-length) between the upper and lower credible limits for each simulation. The relative error of each parameter was estimated as $\text{Bias}_j = \dfrac{1}{K}\sum_{k=1}^{K} |\hat{\theta}_j^k - \theta_j|/\theta_j$, where θ_j is the true parameter value and $\hat{\theta}_j^k$ is the median posterior estimate for the j-th parameter in the k-th Monte Carlo simulation.

Table 3 and Figure 1 show the results for the spatial dependence parameters considering $n = 100$ under different scenarios. The simulation results corresponding to the regression parameters and additional n values for both scenarios can be found in Appendix B1. Since the model in **Scenario 1** has only the intercept, the independent Jeffreys prior is not valid once it provides an improper posterior (Theorem 2.4). Looking over the results for β (see Appendix B1), it can be seen that there is a negligible bias for this parameter except for some atypical points for both the constant and the linear trend. The variability also looks similar, with β_1 and β_2 being the parameters with the higher variance for **Scenario 2**. Regarding credibility intervals, the *reference* and *vague* priors seem to provide closer coverage probability to the nominal value than the *Jeffreys* priors.

On the other hand, the variance parameters seem to present different properties overall priors considered. Looking at the relative error, the *reference* prior seems to provide lower bias than the *vague* and *Jeffreys* priors for σ^2 and ν in both scenarios, this situation is especially remarkable on σ^2, where for each n we can see atypical points which can reach up to 2000% of error for *Jeffreys* priors, this situation is also observed for *reference* and *vague* priors but in a smaller proportion. For the range parameter ϕ, we can see that there is almost no difference in bias for the priors considered in **Scenario 1**, but it changes for the linear trend, where the *reference* and the *vague* priors presented smaller bias.

Regarding credibility intervals properties, it is clear that the proposed *reference* prior is the one that provides the best results. Despite this prior provides wider credibility intervals, its coverage probabilities are more appropriate compared to the other priors, especially when looking at σ^2, where its performance is considerably higher.

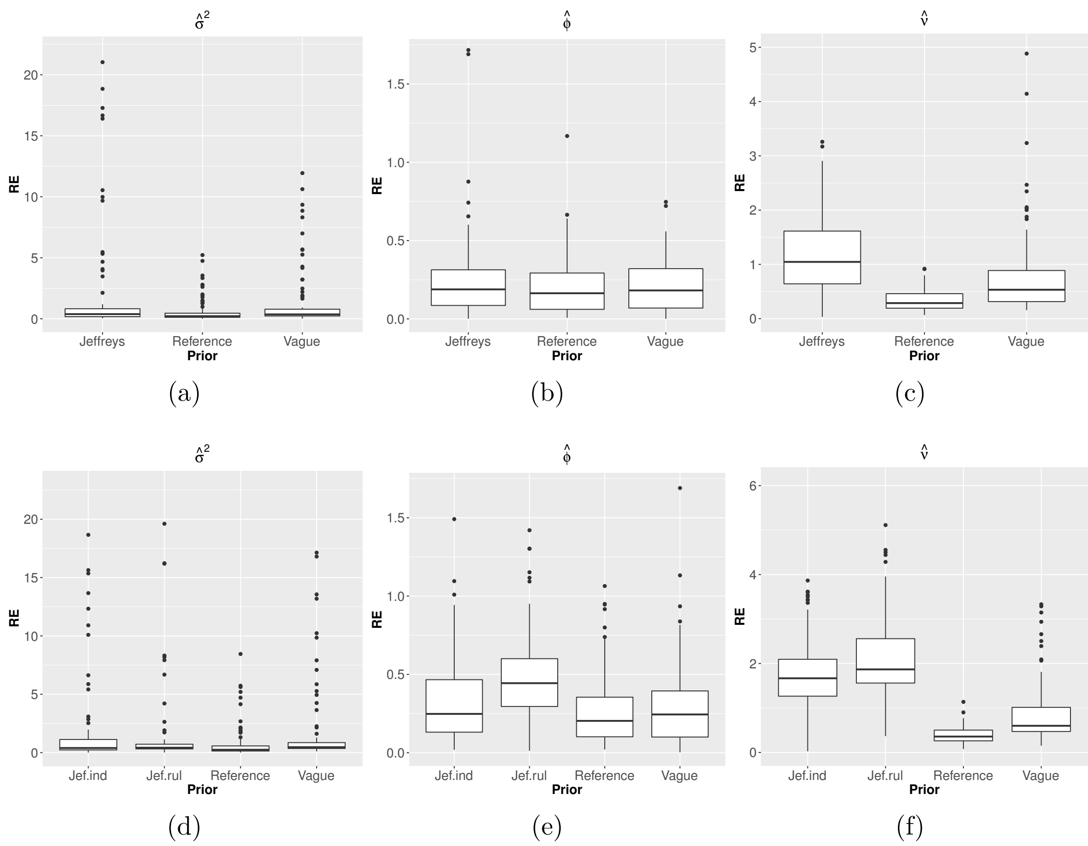


Figure 1 – Simulation study 1. Relative error for the dependence parameters of the T-SR model under Scenario 1 ((a)-(c)) and Scenario 2 ((d)-(f)) for $n = 100$.

3.1.2 Student-t versus Gaussian spatial regression

To compare the performance of the T-SR and Gaussian spatial regression (N-SR) and the different priors, we generated data from a T-SR with the same specifications of the previous Subsection 3.1.1. Then, the T-SR and N-SR models were fitted, using the proposed priors, for the simulated data. The main goal here is to establish if there is any difference between models on both goodnesses of fit and prediction capacity between models and prior choices.

Tables 4 and 5 show information about the relative error and coverage probabilities of the dependence parameters for $n = 100$. Because σ^2 is a scale parameter in the T-SR and a variance parameter in the N-SR model, to make a fair comparison of $\hat{\sigma}^2$, in the N-SR model we compare it to the T-SR variance given by $\frac{\nu}{\nu-2}\sigma^2$. Results for the regression parameters and other sample sizes can be found in Appendix B2. As observed in Subsection 3.1.1, there is negligible bias when looking at the regression parameters for both normal and Student-t cases. Again the estimates under the *Jef rul* provide smaller coverage probability for these parameters in comparison with other priors taken into consideration.

Regarding the spatial dependence parameters, it can be seen that the T-SR recovers σ^2 better than the N-SR having less bias (and standard deviation) for both scenarios and the *reference* and *vague* priors. Regarding the *Jef rul* prior, it appears to have a larger bias for the T-SR case. The range parameter ϕ also appears to have a larger bias for the normal case for all priors, but it is not remarkable.

It can be observed for σ^2 that the coverage probabilities for the *reference* prior is considerably closer to the nominal credibility level 95% for the T-SR and N-SR models in comparison to the other priors. Concerning the range parameter ϕ, the credibility levels look similar for both scenarios, but for $n = 50$ and 100, the expected log-length is smaller for the T-SR case. The N-SR model generates wider intervals for this parameter when either the sample size is small or the data present heavy tail behavior.

To analyze the prediction features, coordinates equivalent to 10% of the original sample size were removed, thus, the total simulated datasets consisted of $n = n_{obs} + n_{pred}$ observations, with $n_{obs} = 45, 90, 225, 450$ the number of observations used to estimate the model (training data) and $n_{pred} = 10\%n$ the number of observations used to assess prediction capacity for both models (test data). Figure 2 shows the MSPE and the expected log-length of each prediction point for both scenarios and $n = 100$, results corresponding to $n = 50, 250, 500$ are in the Appendix B2.

The pointwise predictions are very similar between models under both scenarios, even under model misspecification. This finding is in agreement with Kaufman & Shaby (2013) who state that the mean predictor under model misspecification can be consistent under relatively weak conditions. However, although the log-length of the predictions credible interval looks similar between priors, they do not between processes. From Figure 2 (b) and (d) we can see that the N-SR generates wider intervals. The T-SR credible intervals are, on average, 50.3% smaller than the N-SR one for Scenarios 1 and 2, with similar coverage probabilities.

Table 4 – Simulation study 2. Expected relative error (and standard deviation) of the variance parameters for both Student-t and normal distributions considering $n = 100$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.50 (0.83)	0.21 (0.2)	0.33 (0.19)
		vague	1.24 (2.4)	0.21 (0.18)	0.95 (1.14)
		Jef rul	2.07 (5.68)	0.24 (0.22)	1.17 (0.62)
	N-SR	*reference*	0.84 (1.27)	0.26 (0.21)	-
		vague	1.42 (2.33)	0.32 (0.28)	-
		Jef rul	3.47 (8.01)	0.25 (0.21)	-
Scenario 2	T-SR	*reference*	0.67 (1.26)	0.27 (0.22)	0.42 (0.38)
		vague	1.82 (4.5)	0.26 (0.32)	4.34 (7.3)
		Jef rul	2.71 (8.04)	0.52 (0.36)	2.17 (1.02)
		Jef ind	3.70 (10.8)	0.39 (0.29)	1.90 (0.93)
	N-SR	*reference*	1.25 (1.43)	0.28 (0.17)	-
		vague	1.89 (2.74)	0.46 (0.18)	-
		Jef rul	2.88 (7.28)	0.45 (0.34)	-
		Jef ind	4.64 (9.17)	0.33 (0.25)	-

Table 5 – Simulation study 2. Coverage probability (and expected log-length) of the variance parameters for both Student-t and normal distributions considering $n = 100$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.97 (0.89)	0.98 (0.98)	1 (2.55)
		vague	0.38 (-0.12)	0.85 (0.83)	0.83 (2.43)
		Jef rul	0.44 (0.33)	0.84 (0.76)	0.92 (2.57)
	N-SR	*reference*	0.96 (1.35)	0.93 (1.18)	-
		vague	0.33 (0.03)	0.98 (1.40)	-
		Jef rul	0.49 (0.60)	0.87 (0.98)	-
Scenario 2	T-SR	*reference*	0.95 (0.91)	0.95 (1.08)	1 (2.35)
		vague	0.32 (-0.27)	0.85 (0.9)	0.15 (2.34)
		Jef rul	0.27 (-0.49)	0.61 (0.51)	0 (2.45)
		Jef ind	0.39 (0.33)	0.84 (0.33)	0.09 (0.33)
	N-SR	*reference*	0.99 (1.51)	0.96 (1.32)	-
		vague	0.61 (0.34)	0.97 (1.66)	-
		Jef rul	0.42 (0.02)	0.47 (0.38)	-
		Jef ind	0.60 (0.85)	0.99 (1.50)	-

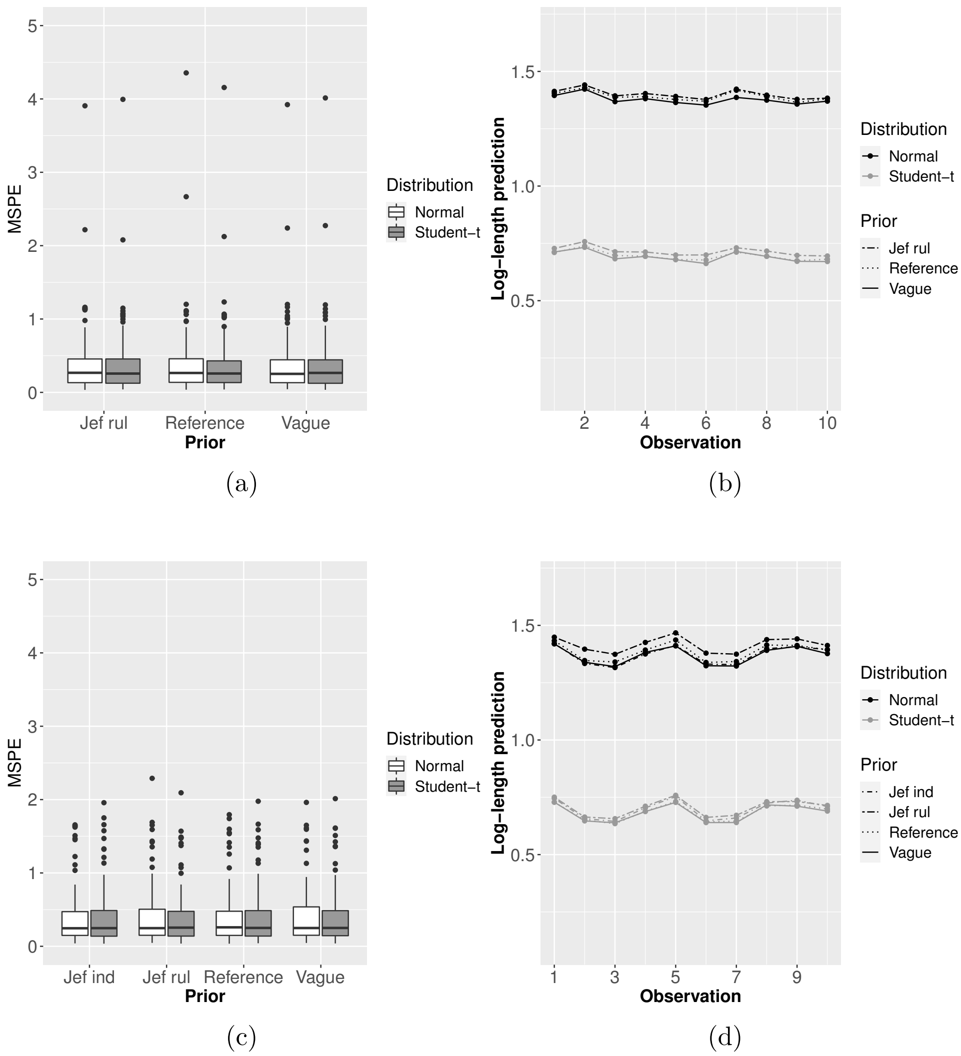
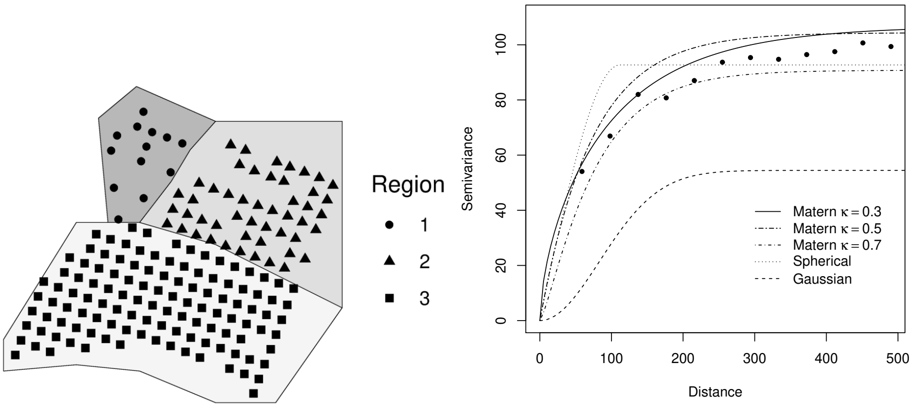


Figure 2 – Simulation study 2. MSPE and log-length of prediction points under Scenario 1 ((a)-(b)) and Scenario 2 ((c)-(d)) for $n = 100$.

3.2 Application

In this section, we illustrate the performance of our reference prior applying it to the calcium content in soil samples data (Ribeiro Jr *et al.*, 2020) (Oliveira, 2003). We use the posterior probability and MSPE to compare different priors and covariance structures. The data was collected by researchers from PESAGRO and EMBRAPA-Solos at the Experimental Campos Station, Rio de Janeiro, Brazil and contains the calcium content measured in soil samples taken from $n = 178$ locations separated at every 50 meters (see Figure 3 (a)). The study area was divided into three regions represented in Figure 3 (a) by different symbols and colors. The first region represent calcium levels in its natural content since it is typically flooded during the rain season and not used as an experimental area. The second region is typically occupied by rice fields and has received fertilizers some time ago. Finally, the third region is an experimental area that has recently received fertilizers. For each sample location, information about elevation is also available. Regarding the correlation structure, Figure 3 (b) shows the variogram for the calcium content dataset. As a first sight, it seems that the Matern family seems to be more adequate for modelling this data, in comparison to other common families of correlation functions.



(a) (b)

Figure 3 – Simulation study 2. MSPE and log-length of prediction points under Scenario 1 ((a)-(b)) and Scenario 2 ((c)-(d)) for $n = 100$.

To study and investigate the different prior proposals, we consider the following T-SR model:

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{x_1} + \beta_1 \mathbf{x_2} + \beta_3 \mathbf{x_3} + \varepsilon,$$

where $\mathbf{Y}$ is the calcium content, $\mathbf{x_1}$ is the elevation, $\mathbf{x_i}$, $i = 2, 3$ are dummy variables indicating the sub-area that the location belongs (1: if y_j belongs to sub-area i, 0: otherwise, $j = 1, \ldots, n$) and $\varepsilon \sim t_n(\mathbf{0}, \sigma^2\mathbf{R}, \nu)$. For the dependence structure, $\mathbf{R}$, we consider three options from the Matérn family (1.3) with $\kappa = 0.3, \kappa = 0.5, \kappa = 0.7$. To adequate the *vague* prior to the characteristics of the calcium data, the prior $\pi(\phi) = U(12.5, 430)$ was selected, such that, it allows the spatial empirical range to vary from the minimum to the maximum distances in the dataset. The model is fitted using the `OBASpatial` package.

Table 6 shows the results for all fitted models under the four different priors. It is important to notice that since our second covariate is categorical, in this case, an intercept can be included to represent the baseline value and the *Jef ind* prior will still provide a proper posterior since this model is equivalent to a model without intercept and 3 levels for the categorical variable. Likewise in the simulation studies the fixed effect parameters seems to be closely estimated by all priors. By looking at the posterior probability $p(m_k|\mathbf{y})$ we can see that the *reference* prior is preferable under all κ's priors and that the *Jef rul* prior presents the lower posterior probability for most of the scenarios. Moreover, under the MSPE criterion, the *reference* prior provides the smaller prediction

Table 6 – Soil samples data. Bayesian estimates, posterior probabilities and MSPE of the models considered of the calcium dataset.

Model	Prior	β_0	β_1	β_2	β_3	σ^2	ϕ	ν	$p(m_k\|\mathbf{y})$	$MSPE$
κ=0.3	*reference*	37.41	1.07	5.75	11.30	119.54	211.27	7.33	0.74	77.16
	vague	38.34	0.37	7.03	13.85	219.98	72.38	9.03	0.04	87.34
	Jef rul	38.98	0.62	6.42	12.66	138.27	175.82	8.93	0.01	84.78
	Jef ind	36.43	0.99	5.87	11.70	125.40	220.13	11.66	0.00	80.56
κ=0.5	*reference*	38.08	1.01	5.47	11.35	135.28	109.49	6.91	0.14	80.93
	vague	39.39	0.01	7.62	14.63	181.01	45.85	9.51	0.01	97.46
	Jef rul	38.73	0.64	6.30	12.81	198.03	97.52	6.98	0.00	88.24
	Jef ind	35.74	1.34	5.45	10.89	192.72	191.77	10.93	0.00	87.42
κ=0.7	*reference*	39.43	0.66	5.35	11.68	110.46	63.26	7.66	0.06	85.64
	vague	39.73	0.05	7.23	15.02	266.98	20.89	6.33	0.00	99.16
	Jef rul	38.18	1.26	3.63	9.78	297.60	153.19	10.61	0.00	93.58
	Jef ind	37.88	1.33	3.90	9.75	245.41	144.03	10.75	0.00	87.82

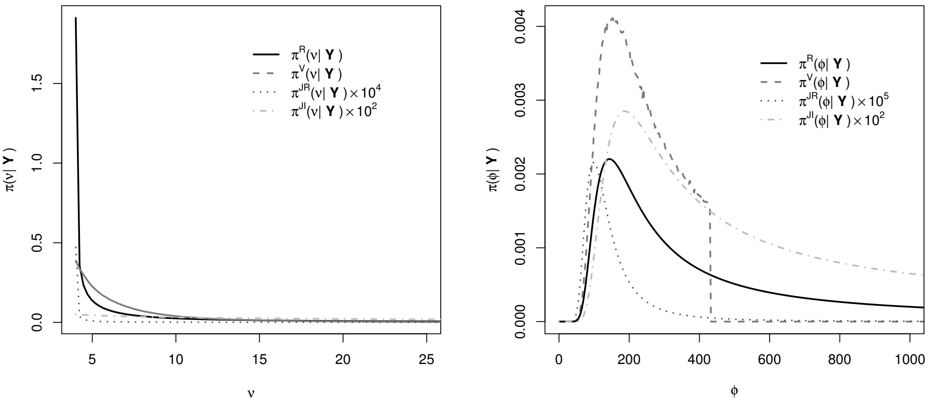
errors followed by the *vague* prior. The best fitting results was provided for the *reference* prior with $\kappa = 0.3$ with higher posterior probability and better predictive power.

For the hyperparameters θ^*, we proceed with our analysis only for $\kappa = 0.3$ (the best model). For all priors the data appears to present a heavy tail behavior with a small value estimated for ν, indicating a lack of adequacy of the N-SR (not fitted). We notice that the *Jef rul* prior is the one that provided the higher estimates for ν. From Figure 4(a), we can see that the posterior for ν have the same shape for all priors, but it is more concentrated for lower values for the *reference* prior and less for the *Jef ind*. Further, the *reference* and *Jeffreys* priors seems to provide similar results for σ^2 and ϕ, while the *vague* prior provides different results. Figure 4(b) shows the posterior distribution for ϕ. From this figure we can see that such difference could be due to the sensitivity of the *vague* prior to the choice of the bounds over ϕ (Berger *et al.*, 2001). Using the estimated ϕ for the *reference* prior, we calculate the spatial range of the process as 418 meters which is 37% the maximum distance, indicating a strong spatial dependence of the calcium content in the soil.

Finally, from this application we can conclude that the introduced *reference* prior presents more stable results for the hyperparameters without affecting its capacity to estimate the fixed effects. Moreover, both model selection criteria show better performance of the reference prior in terms of fit and prediction power. Like observed by in the Gaussian case, the *Jeffreys* priors seem perform poorly with low posterior probability and high MSPE.

3.3 Discussion

In this chapter we propose a *reference* prior for the spatial Student-t regression. Also, two *Jeffery's* type prior were introduced and analyzed. For all the proposed priors,



(a) (b)

Figure 4 – Soil samples data. Marginal posterior distributions for: (a) the degrees of freedom ν and (b) the range parameter ϕ when using the *reference* (R), *vague* (V), *Jef rul* (JR) and *Jef ind* (JI) prior distributions for the soil samples data.

the conditions under which those priors yield a proper posterior distribution were presented and discussed.

We show through simulations that the *reference* prior is the one that presents better performance. Basically, it shows small estimation bias and adequate frequentist coverage for all parameters. Looking for the hyperparameters of interest, out of the other priors choice, the *Jef ind* is the one that provides better C.P., however it has a very strong restriction of not allowing an intercept in the model, making it not suitable for many applications since without an intercept in the model the other fixed effects estimates can be severely biased. In a second simulation study, a comparison between the T-SR and the N-SR and the proposed priors are discussed. When the data present outliers it is possible to see that the *reference* prior provides better coverage for the parameters for both processes. Although the pointwise prediction is very similar between the two models, the T-SR presents smaller prediction credible intervals providing, therefore, more accurate interval estimates with adequate nominal coverage. An illustration with calcium content in the soil was studied to show the applicability and robustness of the proposed reference prior. An `R` package `OBASpatial`, is available at `CRAN` for download and allow practitioners to fit the proposed model with the different priors introduced in the manuscript.

Although the proposed reference prior has shown to have a good performance in the parameters estimates it still presents two limitations. One of them is the restriction that the degrees of freedom ν must be greater that 4, this restriction is necessary because as it can be seen in (2.6) we have that $\nu - 4$ is in the denominator of $\mathbf{A}$ which a cause a discontinuity if not imposed. Moreover, this restriction is also necessary to guarantee the existence of the first two moments of multivariate Student-t distribution (Ho *et al.*, 2012), which makes this restriction necessary in any multivariate Student-t regression models. The second one is that our proposal does not incorporate the nugget effect (or measurement error) as a parameter in the model. The inclusion of the nugget effect makes the calculations for the reference prior and its properties not trivial to be obtained and shown, thus, we understand that this topic is out of the scope of our proposal and must be investigated as future work. It is important to emphasize that even for the N-SR there was a gap of 11 years between the first reference prior proposal and reference priors that includes the nugget effect (Berger *et al.*, 2001; Kazianka & Pilz, 2012; Ren *et al.*, 2012).

4 Penalized complexity priors for the skewness parameter of power links: theoretical settings

4.1 Introduction

Both binary and proportional data are often observed in many applications such as, medicine, public health, social sciences, demographic studies, financial and insurance markets and biological species behavior, to name a few. In practice, generalized linear models with symmetrical links (logit and probit, for example) are the most popular approaches for this kind of data. However, these approaches do not always provide an appropriate fits. In the context of binary regression, symmetrical links are particularly inadequate when the probability of the response variable approaches to zero at different rates than it approaches to one (imbalanced data), this can leads to substantial bias in the mean response estimate (Chen *et al.*, 1999). In this chapter, we focus on a proposal of a new family of prior distributions based on the penalized complexity prior (PC prior) proposed by Simpson *et al.* (2017). Based on the underlying concepts of the PC prior we believe that the proposed priors will provide intuitive and interpretable behaviors and can be naturally calibrated, especially for small samples.

Most of the links from the $(0, 1)$ interval to the real line have been introduced in the literature in the form of binary or binomial regression models. Many of them, aim to handle problems of imbalanced data (de la Cruz *et al.*, 2019).

From a frequentist point of view, Lemonte (2017) introduced the family of parametric power link functions that are considered here. This family includes an extra parameter which represents the skewness of its associated distribution. They performed inference via the maximum likelihood (ML) method and proposed residuals and influence diagnostic procedures to deal with outlying observations and other departures from the model assumptions. Later, Lee & Sinha (2019) investigated the identifiability of the binary regression model with skew-probit link in the presence of both a binary and a continuous covariate. As in the power link family, this link has an additional parameter associated with its skewness. They also studied strategies to reduce the bias of the ML estimator, by penalizing the likelihood through methods proposed by Firth (1993).

On the other hand, from the Bayesian perspective, Chen *et al.* (1999) introduced a new asymmetric link for dichotomous binary data. This was done by including a latent variable in the linear predictor with a skewed distribution. With regard to model fitting, Bayesian inference was performed characterizing the propriety of the posterior using both informative and non-informative priors. Afterwards, Chen *et al.* (2001) considered a skewed logit model to analyze binary responses, Bayesian inference was performed proposing informative priors based on historical information. Additionally, conditional prediction ordinates (CPO) and Bayesian latent residuals were used for model comparison and model adequacy, respectively. Subsequently, Bazán *et al.* (2010) reviewed the various asymmetrical links used in binary regression and proposed a unified approach for two skew-probit links proposed in the literature. This time both frequentist and Bayesian inference were used in model fitting, enabling the existence of ML estimators and the posterior distribution in the presence of improper priors. Later, Naranjo *et al.* (2014) suggested the use of the exponential power (EP) family of distributions as a link to binary regression models. They exploited the mixture representation of this distribution along with a data augmentation framework to improve the efficiency of Gibbs sampling algorithms, in both, informative and non-informative scenarios. A Bayesian approach for power links was introduced in Bazán *et al.* (2017). Among other potential proposals, we can mention the beta regression proposed by Ferrari & Cribari-Neto (2004) and some of its extensions (see, for instance, Galvis *et al.*, 2014; Bayes *et al.*, 2017; Flores *et al.*, 2021).

Although many research works have used power links in the context of binary regression, to the best of our knowledge, there are no studies considering them in a more general context. Moreover, the research field of "practical prior specification" has so far not received, much attention (Simpson *et al.*, 2017). Particularly for power links, common choices of priors have been used for the skewness parameter without solid justifications according to the problem under investigation. In this work, we introduce a new prior distribution for the family of power links (Lemonte, 2017) and establish a general expression of the PC priors for the asymmetry parameter. Bayesian inference is performed, via the Hamiltonian Monte Carlo algorithm, to study additional properties of the proposed priors.

By considering different analysis in a Bayesian approach (model selection criteria, distance to compare predictions with observations, predictive measures and residual analysis), we show that the interpretation of the results in the application can be more convenient when using the PC prior than other common priors studied in the previous literature. Additionally, since the approach to find the PC prior is general, it can be extended to other link functions with an extra parameter proposed for binomial or binary regression models.

The chapter is organized as follows. In Section 4.2, we describe the general hierarchical model with binary, binomial or bounded data. In Section 4.3, e present the PC prior in a general way, as well as, derive a general expression for the particular case of the power link family. Further, we discuss some particular cases in which this expression can be used. Section 4.4 presents the common prior choices for the skewness parameters, full conditionals of the model parameters and the intuition on how to properly specify the hyperparameter under the PC prior.

4.2 Preliminaries

In this section, a brief description of the PC prior (Simpson *et al.*, 2017) is presented and some advantages and disadvantages in comparison to other commonly used priors are discussed.

4.2.1 The Penalized Complexity prior

Simpson *et al.* (2017) stated that the penalized complexity (PC) prior was constructed based on four principles, all of them which consider the existence of a baseline model. The first one, is the Occam's razor or parsimony principal, which states that simpler models is preferable if there is not enough support for a more complex one. The second principle, is the measurement of complexity based on the Kullback-Leibler divergence (see Equation (4.1)). This principle establishes a metric to measure the amount of information lost when approximating a more flexible model with the baseline one. The Kullback-Leibler divergence is defined as

$$KLD\left(f_{\zeta}||f_{\zeta_0}\right) = \int_{x\in\mathcal{X}} f_{\zeta}(x)\log f_{\zeta}(x)dx - \int_{x\in\mathcal{X}} f_{\zeta}(x)\log f_{\zeta_0}(x)dx, \tag{4.1}$$

where $\mathcal{X}$ is the domain of x, $f_{\zeta}(x) = f(x|\zeta)$ and $f_{\zeta_0}(x) = f(x|\zeta = \zeta_0)$ represent density functions associated with the baseline model and its flexible extension, respectively. The third concept is the constant rate penalization, which basically penalizes deviations from the base model with a constant decay rate r. And the last principle is the user defined scale, which declares that the user should have an idea of a sensible size of ζ represented by a tuning parameter λ or a property of the model component. These previous statements imply exponential behavior of the PC prior on the $d(\zeta) = \sqrt{2KLD\left(f_{\zeta}||f_{\zeta_0}\right)}$ scale, which in the univariate space is of the form,

$$\pi(\zeta) = \lambda\exp\{-\lambda d(\zeta)\}\left|\frac{\partial d(\zeta)}{\partial\zeta}\right|.$$

Therefore, the tuning parameter λ is selected to control the contraction of the PC prior towards the baseline model. The intuition behind this prior is to choose λ such that the prior probability of observing a model far from the baseline is small. This can be achieved by defining two quantities: 1) a quantile W of the distance between the flexible distribution and the baseline and 2) its associated probability p_W. With this information and the equality

$$P(d(\zeta) > W) = p_W = \exp(-\lambda W), \tag{4.2}$$

we have that $\lambda = -\log\left(p_W\right)/W$.

Considering the previous statements, this prior is characterized as being informative (although possibly weakly), computationally feasible, indifferent to model parameterization and also, avoids the problem of letting prior specifications to a previous research (the use of common priors as the Gaussian or the uniform, for example). Nonetheless, some care must be taken, because the definition of a "baseline model" might be more subjective than objective, which would generate difficulties when constructing this prior. Also, the modeller's role in choosing the level of complexity λ does not seem to be realistic for hierarchical modeling with several types of parameters, since it only has local meaning by construction (Robert & Rousseau, 2017).

4.3 PC prior for power link functions

In this section, we develop the general form of the PC prior for the family of power links. Furthermore, we also describe some specific cases in which this formula can be reduced to a more straightforward expression.

The parameterization in (1.8) may present inferential issues since its moments depend on α (see the generalized logistic distribution in Subsection 4.3.1, for instance). The problem is that this fact causes the mean and the variance to be affected by the inference of this parameter, so that it is not orthogonal between them, making identifiability between the parameters and convergence of the MCMC harder to obtain. An initial choice of $\mu = 0$ and $\sigma = 1$ seems reasonable since the power links are just extensions of the traditional ones by the inclusion of a skewness parameter. However in this setting these values do not solve the problem. For instance, for the generalized logistic case there is still a dependency among $E(X)$, $V(X)$ and α when the parameters are set at $\mu = 0$ and $\sigma = 1$.

In order to alleviate this confounding issue, the PC prior was constructed over the standardized version of a link function (van Niekerk & Rue, 2020), such that, $E_{f_\alpha}(X) = 0$ and $V_{f_\alpha}(X) = 1$. Let X be a random variable with density:

$$f_\alpha(x) = \frac{\alpha}{\sigma_\alpha} F(x_\alpha)^{\alpha-1} f(x_\alpha), \tag{4.3}$$

where $x_\alpha = {}^{(x-\mu_\alpha)}/_{\sigma_\alpha}$ and μ_α, σ_α are functions of α to guarantee that $E_{f_\alpha}(X) = 0$ and $V_{f_\alpha}(X) = 1$. Then, we say $X \sim PD(\mu_\alpha, \sigma_\alpha, \alpha)$. This parameterization guarantees that inferences about these moments will not be affected by the skewness parameter, making them orthogonal to α. Nonetheless, as in the skew normal case (see, van Niekerk & Rue, 2020; Lee & Sinha, 2019), identifiability problems still persist when an intercept is included in the linear predictor. The next theorem gives a general form for the PC prior under the power distributions. As a particular case of equation (4.3), when $\alpha = 1$ we get $f_1(x) = {}^1/_{\sigma_1} f(x_1)$, which is the density we consider as the baseline link.

Theorem 4.1. *Consider* $f_1(x) = {}^1\!/\!_{\sigma_1} f(x_1)$ *as the baseline model distribution,* $X_\alpha = \dfrac{X - \mu_\alpha}{\sigma_\alpha}$ *and* $KLD(\alpha) = KLD\left(f_\alpha || f_1\right)$*. The PC prior for the power link family can be written as in (4.1), with*

$$\begin{aligned} KLD(\alpha) &= \log\left(\frac{\sigma_1}{\sigma_\alpha}\alpha\right) - \frac{\alpha-1}{\alpha} + E_{f_\alpha}\left\{\log\frac{f(X_\alpha)}{f(X_1)}\right\}, && (4.4)\\ \frac{\partial d(\alpha)}{\partial\alpha} &= \frac{\sqrt{2}}{2}\left\{KLD(\alpha)\right\}^{-\frac{1}{2}}\frac{\partial KLD(\alpha)}{\partial\alpha}, && (4.5)\\ \frac{\partial KLD(\alpha)}{\partial\alpha} &= \frac{\alpha-1}{\alpha^2} - \frac{1}{\sigma_\alpha}\frac{\partial\sigma_\alpha}{\partial\alpha} + \frac{\partial}{\partial\alpha}E_{f_\alpha}\left\{\log\frac{f(X_\alpha)}{f(X_1)}\right\}. && (4.6)\end{aligned}$$

Proof. For the family of links in Equation (1.8), we have,

$$\begin{aligned}\int f_\alpha(x)\log f_\alpha(x)dx &= \log\left(\frac{\alpha}{\sigma_\alpha}\right) + (\alpha-1)\int\frac{\alpha}{\sigma_\alpha}F^{\alpha-1}(x_\alpha)f(x_\alpha)\log F(x_\alpha)dx + \\ &\int f_\alpha(x)\log f(x_\alpha)dx.\end{aligned}$$

Using the change of variable $x_\alpha = {}^{(x-\mu_\alpha)}\!/\!_{\sigma_\alpha}$ we have that,

$$\int\frac{\alpha}{\sigma_\alpha}F^{\alpha-1}(x_\alpha)f(x_\alpha)\log F(x_\alpha)dx = \alpha\int F^{\alpha-1}(x_\alpha)f(x_\alpha)\log F(x_\alpha)dx_\alpha.$$

Let $u = F(x_\alpha)$ and $du = f(x_\alpha)dx$, note that $u \in (0,1)$ given that $f(x_\alpha)$ is also a density function. Therefore,

$$\int F^{\alpha-1}(x_\alpha)f(x_\alpha)\log F(x_\alpha)dx = \int_0^1 u^{\alpha-1}\log(u)du, \tag{4.7}$$

taking $v = \log(u)$ and $dw = u^{\alpha-1}$ we can solve the integral of the right side of Equation (4.7) using integration by parts, that is,

$$\int_0^1 u^{\alpha-1}\log(u)du = vw - \int wdv = \frac{u^\alpha}{\alpha}\log(u) - \frac{u^\alpha}{\alpha^2}\Big|_0^1,$$

also, it can be shown that:

$$\lim_{u\to 0}u^\alpha\log(u) = 0. \tag{4.8}$$

So, using (4.8), we have

$$\int_0^1 u^{\alpha-1}\log(u)du = \frac{u^\alpha}{\alpha}\log(u) - \frac{u^\alpha}{\alpha^2}\Big|_0^1 = -\frac{1}{\alpha^2} - 0 = -\frac{1}{\alpha^2},$$

and

$$\int\frac{\alpha}{\sigma_\alpha}F^{\alpha-1}(x_\alpha)f(x_\alpha)\log F(x_\alpha)dx = -\frac{1}{\alpha},$$

therefore,

$$\int f_\alpha(x)\log f_\alpha(x)dx = \log\left(\frac{\alpha}{\sigma_\alpha}\right) - \frac{\alpha-1}{\alpha} + E_{f_\alpha}\{\log f(X_\alpha)\}. \tag{4.9}$$

Now,

$$\int f_\alpha(x) \log f_1(x) dx = \log\left(\frac{1}{\sigma_1}\right) + E_{f_\alpha}\{\log f(X_1)\}, \tag{4.10}$$

then, substituing (4.9) and (4.10) in Equation (4.1) with $f_\zeta(x) = f_\alpha(x)$ and $f_{\zeta_0}(x) = f_1(x)$ we have

$$KLD(\alpha) = \log\left(\frac{\sigma_1}{\sigma_\alpha}\alpha\right) - \frac{\alpha-1}{\alpha} + E_{f_\alpha}\left\{\log\frac{f(X_\alpha)}{f(X_1)}\right\}. \tag{4.11}$$

To obtain $\pi(\alpha)$, we can rewrite ${}^{\partial d(\alpha)}/_{\partial\alpha}$ as:

$$\frac{\partial d(\alpha)}{\partial\alpha} = \frac{\sqrt{2}}{2}(KLD(\alpha))^{-\frac{1}{2}}\frac{\partial KLD(\alpha)}{\partial\alpha},$$

where,

$$\frac{\partial KLD(\alpha)}{\partial\alpha} = \frac{\alpha-1}{\alpha^2} - \frac{1}{\sigma_\alpha}\frac{\partial\sigma_\alpha}{\partial\alpha} + \frac{\partial}{\partial\alpha}E_{f_\alpha}\left\{\log\frac{f(X_\alpha)}{f(X_1)}\right\}. \tag{4.12}$$

□

Using the stochastic decomposition of $X = \sigma_\alpha Z + \mu_\alpha$ where $Z \sim PD(0,1,\alpha)$ and the fact that $E_{f_\alpha}(X) = 0$ and $V_{f_\alpha}(X) = 1$, we can obtain μ_α and σ_α by solving:

$$\begin{aligned} 0 &= \sigma_\alpha E(Z) + \mu_\alpha, \\ 1 &= \sigma_\alpha^2 V(Z). \end{aligned} \tag{4.13}$$

From this equation system, we have that $\mu_\alpha = -{}^{E(Z)}/_{\sqrt{V(Z)}}$ and $\sigma_\alpha = {}^{1}/_{\sqrt{V(Z)}}$ where $E(Z)$ and $V(Z)$ correspond to the expectation and variance values from the $PD(0,1,\alpha)$ distribution which can be easily obtained using numerical integration. Using the solutions of (4.13), we have the following general formulas:

$$\frac{\partial\mu_\alpha}{\partial\alpha} = \frac{\frac{1}{2\sqrt{V(Z)}}\frac{\partial V(Z)}{\partial\alpha}E(Z) - \sqrt{V(Z)}\frac{\partial E(Z)}{\partial\alpha}}{V(Z)}, \tag{4.14}$$

$$\frac{\partial\sigma_\alpha}{\partial\alpha} = -\frac{1}{2}(V(Z))^{-\frac{3}{2}}\frac{\partial V(Z)}{\partial\alpha}. \tag{4.15}$$

Equations (4.4) - (4.6) can be written in a simpler way for some specific cases (for instance, the power normal distribution). On the other hand, equations (4.14) and (4.15) must be used in those cases where no explicit expressions for μ_α and σ_α are available. The next subsections illustrate particular cases in which Equations (4.4)-(4.6) can be simplified.

4.3.1 Generalized logistic distribution

The random variable X is said to follow a generalized logistic distribution of type I ($X \sim GLI(\mu,\sigma,\alpha)$), if the pdf is given by,

$$f(x|\mu,\sigma,\alpha) = \frac{\alpha}{\sigma}\frac{\exp(-\frac{x-\mu}{\sigma})}{\{1+\exp(-\frac{x-\mu}{\sigma})\}^{\alpha+1}},$$

where μ and σ are location and scale parameters, respectively. For $Z \sim GLI(0,1,\alpha)$ we have that,

$$E(Z) = \psi(\alpha) - \psi(1) \quad V(Z) = \psi'(\alpha) + \psi'(1),$$

where $\psi(a)$ and $\psi'(a)$ are the digamma and trigamma functions, respectively (Gupta & Kundu, 2010). It can be shown that, if $X \sim GLI(\mu,\sigma,\alpha)$ then $Z = {}^{(X-\mu)}/_{\sigma} \sim GLI(0,1,\alpha)$, and

$$E(X) = \mu + \sigma\left(\psi(\alpha) - \psi(1)\right), \quad V(X) = \sigma^2(\psi'(\alpha) + \psi'(1)),$$

so, by setting $E(X) = 0$, $V(X) = 1$ and solving the two equations, we have

$$\mu_\alpha = -\frac{(\psi(\alpha) - \psi(1))}{\sqrt{\psi'(\alpha) + \psi'(1)}} \quad \sigma_\alpha = \sqrt{\frac{1}{\psi'(\alpha) + \psi'(1)}}.$$

Now, the standardized version of the original variable is $X_\alpha = {}^{(X-\mu_\alpha)}/_{\sigma_\alpha}$ and it is used to avoid identifiability problems.

4.3.1.1 PC prior for the GLI distribution

Using the standardized variable X_α, we have that the distribution for the candidate model is given by,

$$f_\alpha(x) = \frac{\alpha}{\sigma_\alpha} \frac{\exp(-x_\alpha)}{(1 + \exp(-x_\alpha))^{\alpha+1}}. \tag{4.16}$$

As a particular case of (4.16), for $\alpha = 1$, we define the baseline model distribution with $\mu_1 = 0$ and $\sigma_1 = \sqrt{{}^{1}/_{2\psi'(1)}}$. The following result gives specific expressions for this link as a consequence of the expressions given in Equations (4.4) - (4.6).

Corollary 1. *The PC prior for the GLI distribution can be written as in Equations (4.4) - (4.6), with*

$$
\begin{aligned}
KLD(\alpha) &= \log\left(\frac{\sigma_1}{\sigma_\alpha}\alpha\right) - \frac{\alpha+1}{\alpha} + \frac{\mu_\alpha}{\sigma_\alpha} + 2E_{f_\alpha}\left[\log\left\{1 + \exp\left(-\frac{X}{\sigma_1}\right)\right\}\right], \\
\frac{\partial d(\alpha)}{\partial \alpha} &= \frac{\sqrt{2}}{2}\left(KLD(\alpha)\right)^{-\frac{1}{2}} \frac{\partial KLD(\alpha)}{\partial \alpha}, \\
\frac{\partial KLD(\alpha)}{\partial \alpha} &= \frac{1}{\alpha\sigma_\alpha}\left(\sigma_\alpha - \frac{\partial \sigma_\alpha}{\partial \alpha}\alpha\right) + \frac{1}{\alpha^2} + \frac{1}{\sigma_\alpha^2}\left(\sigma_\alpha \frac{\partial \mu_\alpha}{\partial \alpha} - \mu_\alpha \frac{\partial \sigma_\alpha}{\partial \alpha}\right) \\
&\quad + 2\frac{\partial}{\partial \alpha} E_{f_\alpha}\left[\log\left\{1 + \exp\left(-\frac{X}{\sigma_1}\right)\right\}\right],
\end{aligned}
$$

with,

$$
\begin{aligned}
\frac{\partial \mu_\alpha}{\partial \alpha} &= -\sigma_\alpha\left(\psi'(\alpha) + \frac{\psi''(\alpha)\mu_\alpha\sigma_\alpha}{2}\right), \\
\frac{\partial \sigma_\alpha}{\partial \alpha} &= \frac{\psi''(\alpha)}{\sigma_\alpha}.
\end{aligned}
$$

Proof. We have that,

$$\begin{aligned} E_{f_\alpha}\{\log f(X_\alpha)\} &= \frac{\mu_\alpha}{\sigma_\alpha} - 2E_{f_\alpha}\left[\log\left\{1+\exp\left(-\frac{X-\mu_\alpha}{\sigma_\alpha}\right)\right\}\right], \\ E_{f_\alpha}\{\log f(X_1)\} &= -2E_{f_\alpha}\left[\log\left\{1+\exp\left(-\frac{X}{\sigma_1}\right)\right\}\right]. \end{aligned}$$

Using the change of variables $u = 1+\exp\left(-\frac{x-\mu_\alpha}{\sigma_\alpha}\right)$ and $du = {}^{-1}/_{\sigma_\alpha}\exp\left(\frac{x-\mu_\alpha}{\sigma_\alpha}\right)dx$,

$$\int f_\alpha(x)\log\left\{1+\exp\left(-\frac{x-\mu_\alpha}{\sigma_\alpha}\right)\right\}dx = \alpha\int\frac{\log(u)}{u^{\alpha+1}}du = -\alpha\int u^{-(\alpha+1)}\log(u)du.$$

We can solve the last integral by parts by taking $v = \log(u)$ and $dw = u^{-(\alpha+1)}$. In this particular case $-\infty < x < \infty$, so, $1 < u < \infty$. Also, we can show that:

$$\lim_{u\to\infty} u^{-\alpha}\log(u) = 0.$$

Using these facts jointly, we have:

$$\begin{aligned} \int f_\alpha(x)\log\left\{1+\exp\left(-\frac{x-\mu_\alpha}{\sigma_\alpha}\right)\right\}dx &= \alpha\left(\frac{-u^{-\alpha}}{-\alpha}\log(u) - \frac{u^{-\alpha}}{\alpha^2}\right)\Big|_1^\infty \\ &= -\frac{\alpha}{\alpha^2} - 0 = -\frac{1}{\alpha}, \end{aligned}$$

then,

$$E_{f_\alpha}\left[\log\left\{\frac{f(X_\alpha)}{f(X_1)}\right\}\right] = \frac{\mu_\alpha}{\sigma_\alpha} + \frac{2}{\alpha} + 2E_{f_\alpha}\left[\log\left\{1+\exp\left(-\frac{X}{\sigma_1}\right)\right\}\right].$$

Finally, using (4.4), we have that the Kullback-Leibler divergence for this distribution is:

$$KLD(\alpha) = \log\left(\frac{\sigma_1}{\sigma_\alpha}\alpha\right) - \frac{\alpha+1}{\alpha} + \frac{\mu_\alpha}{\sigma_\alpha} + 2E_{f_\alpha}\left[\log\left\{1+\exp\left(-\frac{X}{\sigma_1}\right)\right\}\right],$$

with

$$\begin{aligned} \frac{\partial KLD(\alpha)}{\partial\alpha} &= \frac{1}{\alpha\sigma_\alpha}\left(\sigma_\alpha - \frac{\partial\sigma_\alpha}{\partial\alpha}\alpha\right) + \frac{1}{\alpha^2} + \frac{1}{\sigma_\alpha^2}\left(\sigma_\alpha\frac{\partial\mu_\alpha}{\partial\alpha} - \mu_\alpha\frac{\partial\sigma_\alpha}{\partial\alpha}\right) \\ &\quad +2\frac{\partial}{\partial\alpha}E_{f_\alpha}\left[\log\left\{1+\exp\left(-\frac{X}{\sigma_1}\right)\right\}\right]. \end{aligned}$$

□

4.3.2 Power normal distribution

Unfortunately when X follows a power normal (PN) distribution there are not closed expressions for $E(X)$ and $V(X)$. However, we can obtain a simplified version of the PC prior getting μ_α and σ_α through simple numerical integration. Its density function is given by:

$$f_\alpha(x) = \frac{\alpha}{\sigma_\alpha}\Phi(x_\alpha)^{\alpha-1}\phi(x_\alpha),$$

where $x_\alpha = {}^{(x-\mu_\alpha)}/_{\sigma_\alpha}$ and $\phi(.)$,$\Phi(.)$ correspond to the pdf and cdf of the standard normal univariate random variable, respectively. Note that for the baseline model ($\alpha = 1$), we have $\mu_1 = 0$ and $\sigma_1 = 1$ because the analysis of these choices guarantees that $E_{f_\alpha}(X) = 0$ and $V_{f_\alpha}(x) = 1$.

4.3.2.1 PC prior for the power normal distribution

Like the GLI distribution, Equations (4.4) - (4.6) can be used in order to get a simpler version of the PC prior for the PN link. This is summarized in the next corollary

Corollary 2. *The PC prior for the PN distribution can be written as in (4.1), with*

$$\begin{aligned}
KLD(\alpha) &= \log(\alpha) - \frac{\alpha-1}{\alpha} + \frac{1}{2} - \log\sigma_\alpha - \frac{1+\mu_\alpha^2}{2\sigma_\alpha^2}, \\
\frac{\partial d(\alpha)}{\partial\alpha} &= \frac{\sqrt{2}}{2}(KLD(\alpha))^{-\frac{1}{2}}\frac{\partial KLD(\alpha)}{\partial\alpha}, \\
\frac{\partial KLD(\alpha)}{\partial\alpha} &= \frac{\alpha-1}{\alpha^2} + \frac{1}{\sigma_\alpha}\frac{\partial\sigma_\alpha}{\partial\alpha}\left(\frac{1+\mu_\alpha^2}{\sigma_\alpha^2} - 1\right) - \frac{\mu_\alpha}{\sigma_\alpha^2}\frac{\partial\mu_\alpha}{\partial\alpha},
\end{aligned}$$

where ${}^{\partial\mu_\alpha}/_{\partial\alpha}$ *and* ${}^{\partial\sigma_\alpha}/_{\partial\alpha}$ *can be computed as in* (4.14) *and* (4.15).

Proof. Let $X_\alpha = {}^{(X-\mu_\alpha)}/_{\sigma_\alpha}$, by computing the expectation term of Equation (4.4) for the power normal case, we have that:

$$\begin{aligned}
E_{f_\alpha}\{\log f(X_\alpha)\} &= \int \frac{\alpha}{\sigma_\alpha}\Phi(x_\alpha)^{\alpha-1}\phi(x_\alpha)\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}x_\alpha^2\right)dx \\
&= -\frac{1}{2}\log 2\pi - \frac{1}{2}E_{f_\alpha}(X_\alpha^2).
\end{aligned}$$

We already have, by construction, that $E_{f_\alpha}(X) = 0$ and $V_{f_\alpha}(X) = 1$. Therefore, $E_{f_\alpha}(X^2) = 1$ and

$$E_{f_\alpha}(X_\alpha^2) = E_{f_\alpha}\left\{\left(\frac{X-\mu_\alpha}{\sigma_\alpha}\right)^2\right\} = \frac{1+\mu_\alpha^2}{\sigma_\alpha^2},$$

so,

$$E_{f_\alpha}\{\log f(X_\alpha)\} = -\frac{1}{2}\log 2\pi - \frac{1+\mu_\alpha^2}{2\sigma_\alpha^2}. \tag{4.17}$$

Also,

$$\begin{aligned}
E_{f_\alpha}\{\log f(X_1)\} &= \int \frac{\alpha}{\sigma_\alpha}\Phi(x_\alpha)^{\alpha-1}\phi(x_\alpha)\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}x^2\right)dx \\
&= -\frac{1}{2}\log 2\pi - \frac{1}{2}. \qquad (4.18)
\end{aligned}$$

Substituting (4.17) and (4.18) in (4.4), we have that

$$KLD(\alpha) = \log(\alpha) - \frac{\alpha-1}{\alpha} + \frac{1}{2} - \log\sigma_\alpha - \frac{1+\mu_\alpha^2}{2\sigma_\alpha^2}$$

and
$$\frac{\partial KLD(\alpha)}{\partial \alpha} = \frac{\alpha-1}{\alpha^2} + \frac{1}{\sigma_\alpha}\frac{\partial \sigma_\alpha}{\partial \alpha}\left(\frac{1+\mu_\alpha^2}{\sigma_\alpha^2} - 1\right) - \frac{\mu_\alpha}{\sigma_\alpha^2}\frac{\partial \mu_\alpha}{\partial \alpha}.$$

□

4.3.3 Power Student-t distribution

A random variable X follows a power Student-t (PT) distribution if its pdf is given by,

$$f_\alpha(x) = \frac{\alpha}{\sigma_\alpha} T_\nu(x_\alpha)^{\alpha-1} t_\nu(x_\alpha),$$

where t and T correspond to the pdf and cdf of a standard Student-t univariate random variable with ν degrees of freedom. Note that for the base model ($\alpha = 1$) we have $\mu_1 = 0$ and $\sigma_1 = \dfrac{\nu-2}{\nu}$ which assure that $E_{f_\alpha}(X) = 0$ and $V_{f_\alpha}(X) = 1$.

Like in the power normal case, no closed form expressions for $E(X)$ and $V(X)$ can be obtained, or a more straightforward expression in terms of its first two moments. Thus, the PC prior for the power Student-t link, needs to be numerically computed by using the general Equations (4.4)-(4.6) to obtain the $KLD(\alpha)$ measure and its derivative as a function of μ_α and σ_α.

4.3.4 Power Gumbel distribution

A random variable X follows the power Gumbel distribution ($X \sim PG(\mu, \sigma, \alpha)$) if the pdf is given by,

$$f(x|\mu,\sigma,\alpha) = \frac{\alpha}{\sigma}\exp\left\{-\alpha\exp\left(-\frac{x-\mu}{\sigma}\right) - \left(\frac{x-\mu}{\sigma}\right)\right\}. \tag{4.19}$$

Equation (4.19) can be used to obtain the first two moments of this distribution, which are necessary to get the standardized variable X_α. This will be done using the moments generating function of this density.

Proposition 4.1. *If $X \sim PG(0, 1, \alpha)$, then,*

$$m_X(t) = E\{\exp(tX)\} = \alpha^t \Gamma(1-t),$$

and as a consequence, we have that $E(X) = \log(\alpha) - \Gamma'(1)$ and $V(X) = \dfrac{\pi^2}{6}$.

Proof. We have that,

$$\begin{aligned} E\left(\exp(tX)\right) &= \int \exp(tX)\alpha\exp\{-\alpha\exp(-x) - x\}dx \\ &= \alpha\int \exp(tx)\exp(-x)\exp\{-\alpha\exp(-x)\}dx, \end{aligned}$$

now, taking $x = -\log(u)$, $dx = -{}^{du}/{}_{u}$, we have:

$$\begin{aligned}\alpha\int \exp(tx)\exp(-x)\exp\{-\alpha\exp(-x)\}dx &= \alpha\int \exp\{-t\log(u)\}\exp(-\alpha u)du\\ &= \alpha\int u^{-t}\exp(-\alpha u)du\\ &= \alpha\frac{\Gamma(1-t)}{\alpha^{1-t}}\\ &\quad \int\frac{\alpha^{1-t}}{\Gamma(1-t)}u^{(1-t)-1}\exp(-\alpha u)du,\end{aligned}$$

note that since $-\infty < x < \infty$, then $0 < u < \infty$ and therefore, the term inside the last integral corresponds to the gamma distribution. Since this integral is 1, the result follows. As a consequence of Equation (21) of the main paper, we have that:

$$\begin{aligned}m'_X(t) &= \alpha^t\log(\alpha)\Gamma(1-t) - \alpha^t\Gamma'(1-t)\\ m''_X(t) &= \log(\alpha)m'_X(t) + \alpha^t\Gamma''(1-t) - \alpha^t\log(\alpha)\Gamma'(1-t),\end{aligned}$$

and therefore, $E(X) = m'_X(0) = \log(\alpha) - \Gamma'(1)$ and

$$\begin{aligned}E(X^2) = m''_X(0) &= \log(\alpha)\{\log(\alpha) - \Gamma'(1)\} + \Gamma''(1) - \log(\alpha)\Gamma'(1)\\ &= \log^2(\alpha) - 2\log(\alpha)\Gamma'(1) + \Gamma''(1),\end{aligned}$$

then,

$$\begin{aligned}Var(X) &= \log^2(\alpha) - 2\log(\alpha)\Gamma'(1) + \Gamma''(1) - \{\log(\alpha) - \Gamma'(1)\}^2\\ &= \Gamma''(1) - \{\Gamma'(1)\}^2 = \frac{\pi^2}{6}.\end{aligned}$$

□

Considering $Y = \sigma X + \mu \sim PG(\mu, \sigma, \alpha)$, and the restrictions $E(Y) = 0$ and $V(Y) = 1$, we can get the standardized form X_α by setting,

$$\mu_\alpha = -\sqrt{\frac{6}{\pi^2}}(\log(\alpha) - \Gamma'(1)), \text{ and } \sigma_\alpha = \sqrt{\frac{6}{\pi^2}}.$$

Note that σ does not depend on α. Using this fact, Theorem 4.2 shows that this family is non-identifiable for α.

Theorem 4.2. *For the power Gumbel distribution, $f_\alpha(x)$ does not depend on α.*

Proof. Equation (4.19) establishes that the power Gumbel distribution has, in terms of α, $\mu_\alpha = -c_1(\log\alpha - c_2)$ and $\sigma_\alpha = c_1$, where $c_1 = \sqrt{6/\pi^2}$, $c_2 = \Gamma'(1)$. Then, substituting these values in (4.19), we have

$$f_\alpha(x) = \frac{\alpha}{c_1}\exp\left\{-\alpha\exp\left(-\frac{x + c_1(\log\alpha - c_2)}{c_1}\right)\right\}\exp\left\{-\frac{x + c_1(\log\alpha - c_2)}{c_1}\right\}, \quad (4.20)$$

but,

$$\exp\left\{-\frac{x + c_1(\log\alpha - c_2)}{c_1}\right\} = \frac{1}{\alpha}\exp\left(c_2 - \frac{x}{c_1}\right) \tag{4.21}$$

and substituting (4.21) in (4.20), we obtain,

$$f_\alpha(x) = \frac{1}{c_1}\exp\left\{-\exp\left(c_2 - \frac{x}{c_1}\right)\right\}\exp\left(c_2 - \frac{x}{c_1}\right),$$

which does not depend on α, so, the result follows. □

Based on Theorem 4.2, the power Gumbel link proposed by Bazán *et al.* (2017) is not well defined, so, it is senseless to use it as a valid alternative in modeling.

4.4 Bayesian inference

In this section, we present the elements required to perform Bayesian estimation for the model (1.7) using the family of link functions in (1.8). We give a brief description of some of the most common priors used in the literature (which we also use for comparison purposes) and general expressions for the full conditional distributions according to each prior considered. Finally, we state a way of choosing α according to the tail behavior of one of the priors we use in the comparison procedure.

4.4.1 Commonly used priors

Let $\theta = (\boldsymbol{\beta}, \alpha, \boldsymbol{\gamma})$ be the model parameters. The first step is to define the prior distribution $\pi(\boldsymbol{\beta}, \alpha, \boldsymbol{\gamma})$. Therefore, we consider an independent structure where

$$\pi(\boldsymbol{\beta}, \alpha, \gamma) \propto \pi(\boldsymbol{\beta})\pi(\alpha)\pi(\boldsymbol{\gamma}). \tag{4.22}$$

Since $\boldsymbol{\beta}$ is a vector, we consider $\pi(\boldsymbol{\beta}) \sim N_p(\mu_{\boldsymbol{\beta}}, \Sigma_{\boldsymbol{\beta}})$ as prior where $\mu_{\boldsymbol{\beta}}$ is the mean vector and $\Sigma_{\boldsymbol{\beta}}$ the covariance matrix. For the skewness parameter prior $\pi(\alpha)$, Bazán *et al.* (2017) considered the reparameterization $\delta = \log(\alpha)$ and set a uniform prior for δ in such a way that for the uniform prior, $\alpha \in (\exp(-2), \exp(2))$. Given the support of α, another common alternative is to use a truncated normal (TN) distribution. In the skew probit regression context, van Niekerk & Rue (2020) considered a PC prior for α by comparing it with a Gaussian distribution of parameters $\mu = 0$ and $\sigma^2 = 10^2$. Note that for the extreme cases $\alpha \to 0$ and $\alpha \to \infty$, the success probability p_i goes to 1 and 0, respectively, in such a way that the random variables Y_i become degenerate. So, a good way to establish hyperparameters when considering common priors without previous information, is to set them in such a way that these extreme cases are taken into account.

The $\boldsymbol{\gamma}$ vector accommodates any other parameters of the distribution of interest, so an adequate prior specification depends on the likelihood distribution.

4.4.2 Posterior analysis

Without loss of generality, as commonly observed in the literature, we present our analysis for a binomial regression. Thus, $\theta = (\boldsymbol{\beta}, \alpha)$ and we can define the likelihood as

$$L(\theta|\mathbf{y}) = \prod_{i=1}^{n} \binom{n_i}{y_i} p_i^{y_i} (1-p_i)^{n_i - y_i} \tag{4.23}$$

where $p_i = F(\mathbf{x}_i^\top \boldsymbol{\beta})$ is the probability of success and n_i is the number of Bernoulli trials at observation i. Then, the posterior distribution for $\theta = (\boldsymbol{\beta}, \alpha)$ is given by,

$$\pi(\theta|\mathbf{y}) \propto \pi(\theta) L(\theta|\mathbf{y}),$$

where $\pi(\theta)$ has the form (4.22) and $L(\theta|\mathbf{y})$ has the form (4.23). Starting from this equation, we can derive the full conditional distribution for $\boldsymbol{\beta}$ as,

$$\pi(\boldsymbol{\beta}|\alpha, \mathbf{y}) \quad \propto \quad \exp\left(-(\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}})^\top \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} (\boldsymbol{\beta} - \mu_{\boldsymbol{\beta}})\right) \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{n_i - y_i}.$$

For α, the full conditionals depend on the prior choice discussed in the previous subsection. Taking into account the prior reparameterization $\delta \sim U(-b, b)$, we have that $\alpha = \exp(\delta)$. Using this fact, the explicit expressions of the full conditional distributions for the skewness parameter given the priors we consider here, have the form,

$$\begin{aligned}
\pi_{U_\delta}(\alpha|\boldsymbol{\beta}, \mathbf{y}) \quad &\propto \quad \frac{1}{\alpha} \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{n_i - y_i}, \\
\pi_{TN}(\alpha|\boldsymbol{\beta}, \mathbf{y}) \quad &\propto \quad \exp \frac{(\alpha - \mu_{TN})^2}{2\sigma_{TN}^2} 1(\alpha)_{(0,+\infty)} \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{n_i - y_i}, \\
\pi_{PC}(\alpha|\boldsymbol{\beta}, \mathbf{y}) \quad &\propto \quad \lambda \exp\{-\lambda d(\alpha)\} \left| \frac{\partial d(\alpha)}{\partial \alpha} \right| \prod_{i=1}^{n} p_i^{y_i} (1-p_i)^{n_i - y_i},
\end{aligned}$$

where the indexes U_δ, TN and PC denote the resulting posteriors obtained from the uniform δ, the TN and the PC priors respectively, while (μ_{TN}, σ_{TN}) and λ are the hyperparameters associated with the TN and the PC priors and $1(\alpha)_{(0,+\infty)}$ represents the indicator function, according to which α belongs to the positive real line.

Unfortunately, the expressions above do not have closed form, making hard to establish a MCMC scheme to obtain Bayesian inferences. The NUTS method, briefly described in Section 1.1.6, is already implemented in the `STAN` software available in the `R` statistical software (R Core Team, 2020) through the `rstan` package (Stan Development Team, 2020). It is used to perform all of our Bayesian analyses in this chapter.

4.4.3 Choice of λ

The main purpose of this chapter is to introduce a new prior for α. To provide evidence of its advantages, we are interested in assessing its performance in comparison with commonly used ones. As mentioned in Subsection 4.4.1, the uniform (for $\delta = \log(\alpha)$) and TN priors are the most common choices in the literature. To make a fair comparison between the priors, we construct the priors to be comparable with each other. This premisse guides our choices of λ and the others priors parameters, as explained next.

Equation (4.2) presents the criteria which set λ as a result of the tail behavior of $d(\zeta)$ for the PC prior. Using this fact, we fix the tail behavior of one of the priors and select the other prior parameters to mimic it. Since $\alpha > 0$, let us consider the TN prior in $(0, +\infty)$ as our baseline traditional prior. Note that for this distribution,

$$P(\alpha > 2\sigma) = p_{TN}, \tag{4.24}$$

where σ is the scale parameter for this prior. Starting from this point, we can fix $W = d(2\sigma)$ and $p_W = p_{TN}$ in Equation (4.2) to obtain a similar tail for the PC prior. The value of λ will depend on the power link used in our analyses. For the uniform prior this can also be achieved by replacing $\alpha = \exp(\delta)$ and using the fact that $\delta \sim U(-b, b)$, $b > 0$. In this way, one can find b such that Equation (4.24) is satisfied.

5 Penalized complexity priors for the skewness parameter of power links: applications with simulated and real data

In this chapter, a simulation study is reported to assess the frequentist properties of the Bayesian estimates under different priors. In Section 5.2 a real dataset is analyzed to validate our findings. Finally, some closing remarks are discussed in Section 5.3.

5.1 Simulation Study

In this section we present a simulation study to assess the performance of the proposed prior for the skewness parameter α. A total of $m = 200$ binary datasets were simulated using the generalized logit (GL) and the power normal (PN) links. The power robit link (Subsection 4.3.3) requires numerical calculations that we were unable to code in `rstan`. For this reason, this link is not included here. Furthermore, as mentioned in Section 4.3, there is an identifiability issue between the intercept and α. With the goal of correctly studying the prior sensitivity effect on the model fit we opted to not include it in the linear predictor.

To simulate the datasets, we independently generated $x_{1i} \sim N(0, 1^2)$, $x_{2i} \sim N(1, 3^2)$ and $x_{3i} \sim N(0.5, 2^2)$ for $i = 1 \dots n$ taking $\beta = (3, 0.8, 1.7)'$ and $\alpha = 2.2$. Using these results we got the linear predictor $\mathbf{x}_i'\boldsymbol{\beta}$ where $\mathbf{x}_i' = (x_{1i}, x_{2i}, x_{3i})$ and $p_i = F(\mathbf{x}_i'\boldsymbol{\beta})$ for both links in each sample. Finally, we simulated y_i by following a Bernoulli distribution with parameter p_i considering $n = 50, 100$ and 250. The likelihood in this scenario was obtained by setting $n_i = 1$ in Equation (4.23).

Concerning the prior structure for the linear predictor parameters, we considered independent $\beta_j \sim N(0, 100)$ for $j = 1, 2, 3$. For the skewness parameter, to evaluate the performance of the PC prior, we compared it with the TN and the uniform prior (for $\delta = \log(\alpha)$) using different criteria. Following the idea presented in Subsection 4.4.3, we defined the tail behavior for each distribution by setting $\alpha \sim TN(0, 10^2, \mathbb{A})$, where $\mathbb{A} = (0, +\infty)$. The resulting hyperparameters were $\lambda_{GL} = 8.936$ and $\lambda_{PN} = 15.479$ for the PC prior using the GL and PN links respectively, and $b = 3.296$ for the uniform prior.

Regarding the comparison criteria, we calculated 1) the relative error, 2) the expected log-length of the highest posterior 95% credible interval and 3) the coverage probability for each parameter. The first one, was computed as $\text{Err}_j = \dfrac{1}{M}\sum_{m=1}^{M} |\hat{\theta}_j^m - \theta_j|/\theta_j$, where θ_j is the true parameter value and $\hat{\theta}_j^M$ is the mean posterior estimate for the j-th parameter in the m-th Monte Carlo simulation. The second one was calculated as the mean of the logarithm of the difference (log-length) between the upper and lower credible interval limits for each simulation. Finally, the last one was computed as the number of simulations where the true parameter lies inside the credible limits.

Tables 7 and 8 show the log-length and the coverage probability for the credibility intervals of α and $\boldsymbol{\beta}$ considering the GL and PN links. For the credibility intervals for α, it can be observed that, for the GL link, the coverage probability is highest when $n = 50$ and the PC prior is used. For the PN link on the other hand, the coverage probability for the TN prior is slightly higher, but the PC prior shows much smaller interval log-lengths. For $n = 100, 250$, despite the fact the coverage probability seems to reach the theoretical 95% level for all priors, the interval lengths are still much shorter for the PC prior. Although all models have the same prior for $\boldsymbol{\beta}$, the selected prior for α also affects its posterior results. Thus, concerning $\boldsymbol{\beta}$, the estimated intervals generated by the PC prior model under the GL link present better properties in terms of log-length and coverage probability, with β_2 being the best estimated parameter. Under the PN link and for $n = 50$, it can be observed that the model using the PC prior presents larger intervals, although its coverage probabilities are higher compared with the models using the two other priors. For $n = 100, 250$, these properties for the other links improve in relation to the model with the PC prior.

Table 7 – Simulation study. Expected log-length (and coverage probability) of the skewness parameter α with both the GL and PN links.

		Prior		
Link	n	PC	TN	Uniform
GL	50	1.54 (0.85)	2.09 (0.80)	2.05 (0.78)
	100	1.08 (0.90)	1.49 (0.90)	1.42 (0.93)
	250	0.47 (0.95)	0.6 (0.95)	0.56 (0.95)
PN	50	1.68 (0.92)	2.42 (0.95)	2.38 (0.91)
	100	1.49 (0.93)	1.94 (0.92)	1.90 (0.93)
	250	0.75 (0.94)	0.92 (0.94)	0.87 (0.93)

Figure 5 shows the relative error for α. The results for $\boldsymbol{\beta}$ are available in Figures 15-17 (Appendix C1). Under this metric, the model using the PC prior shows better results particularly for the skewness parameter α, where for both links, this measure is considerably smaller compared with the other priors. This behavior is more pronounced when n is small but is also valid for larger ones. Regarding $\boldsymbol{\beta}$, the model using the PC prior under the GL link gives less biased estimates, while for the PN link this is more pronounced for $n = 50$, and for the larger sample sizes similar results between the priors are observed.

Table 8 – Simulation study. Expected log-length (and coverage probability) of the linear predictor parameters with both the GL and PN links.

		GL			PN		
n	$\boldsymbol{\beta}$	PC	TN	Uniform	PC	TN	Uniform
50	β_1	1.95 (0.88)	2.02 (0.72)	2.04 (0.75)	2.18 (0.76)	2.13 (0.66)	2.16 (0.68)
	β_2	0.58 (0.89)	0.66 (0.69)	0.68 (0.74)	0.78 (0.80)	0.75 (0.61)	0.77 (0.69)
	β_3	1.37 (0.84)	1.44 (0.67)	1.46 (0.70)	1.54 (0.74)	1.51 (0.60)	1.52 (0.63)
100	β_1	1.27 (0.90)	1.34 (0.80)	1.33 (0.84)	1.47 (0.87)	1.49 (0.71)	1.50 (0.77)
	β_2	-0.03 (0.90)	0.04 (0.79)	0.04 (0.84)	0.15 (0.84)	0.16 (0.70)	0.18 (0.75)
	β_3	0.63 (0.92)	0.71 (0.78)	0.71 (0.83)	0.88 (0.85)	0.89 (0.70)	0.91 (0.74)
250	β_1	0.60 (0.93)	0.61 (0.90)	0.61 (0.90)	0.79 (0.90)	0.80 (0.85)	0.80 (0.86)
	β_2	-0.69 (0.93)	-0.67 (0.87)	-0.68 (0.87)	-0.56 (0.91)	-0.54 (0.83)	-0.55 (0.85)
	β_3	0.01 (0.94)	0.03 (0.85)	0.02 (0.89)	0.19 (0.90)	0.20 (0.85)	0.20 (0.87)

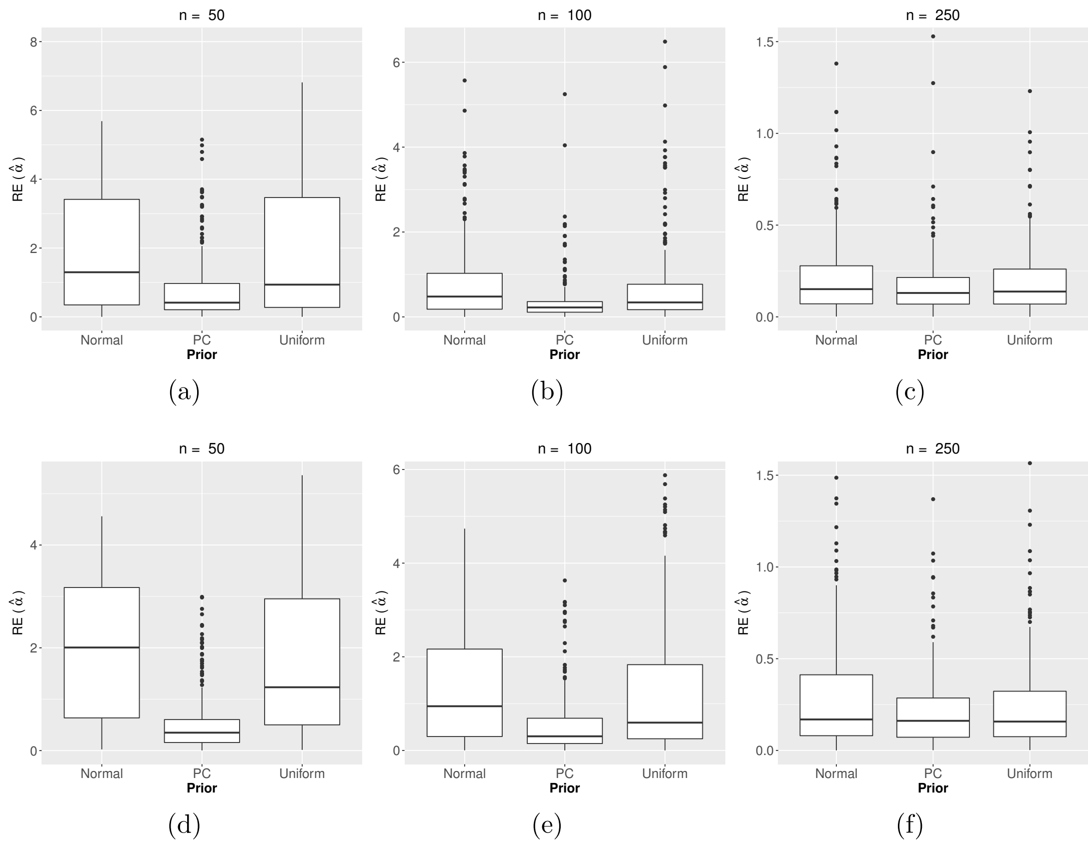


Figure 5 – Simulation study. Relative errors for the skewness parameter α with the GL link ((a)-(c)) and the PN link ((d)-(f)) and different sample sizes.

5.2 Application

In this section we illustrate the good performance of the PC prior for selecting the skewness parameter in a binomial regression model with power links considering a real dataset. Specifically, we propose different analysis using the Bayesian approach for bladder cancer data (Devidas *et al.*, 1993).

This dataset refers to the effects on mice continuously exposed to low doses of 2-acetylaminofluorene (2-AAF). Doses of this biochemical were fed to female mice, from weanling until the mice were either sacrificed, became moribund or died. The dose or concentration of 2-AAF in parts per million (ppm) (x_i) was fixed and the number of mice with tumors (y_i) out of the number of mice exposed (n_i) was registered as shown in data in Table 9, reproduced from Devidas *et al.* (1993).

Table 9 – Observed bladder cancer data for mice exposed to 2-AAF

Dose (ppm)	0	30	35	45	60	75	100	150
No. mice with tumors	1	5	0	2	2	12	21	11
No. mice exposed	101	443	200	103	66	75	31	11

In the binomial regression model, GL and PN links were considered for p_i and a quadratic effect of the doses was tested because some investigators report that the dose-response curve begins to straighten at higher doses, leading some to propose that a linear-quadratic-linear dose response model is more appropriate (Astrahan, 2008). Thus, the linear predictor is $\eta_i = \beta_1 x_i + \beta_2 x_i^2$ where as indicated in Section 4.3, because of identifiability issues, the intercept was no considered in the linear predictor.

With respect to prior specification, we consider an independent Gaussian prior for $\boldsymbol{\beta}$ in all models, i.e., $\beta_i \sim N(0, 100)$ for $i = 1, 2$. For the skewness parameter α on the other hand, we defined the priors as in Section 5.1, i.e., $\alpha \sim TN(0, 10^2, \mathbb{A})$ with $\mathbb{A} = (0, +\infty)$, $\alpha \sim U(-b, b)$ with $b = 3.296$ and $\alpha \sim PC(\lambda)$ with $\lambda_{PL} = 8.936$ and $\lambda_{PN} = 15.479$ for the GL and PN links, respectively. To fit the models, two chains with different initial values were run. For each of them, a total of 5000 posterior samples with no thinning were generated and the first 1000 were discarded as burn-in. Therefore, we ended up with a final posterior sample of 8000 observations for each parameter. The convergence of the HMC was verified by the Gelman & Rubin (1992) criterion. To exemplify the convergence of the posterior sample obtained by the models fitted, Figure 18 (Supplement Material C2) shows the trace and ACF plots for the parameters of the PC link with the PC prior.

Table 10 shows some of the most common model selection criteria (see Subsection 1.1.9) for the bladder cancer data considering PN and GL links under different prior specifications for the parameter α.

Table 10 – Bladder cancer data. Model selection criteria for the GL and PN links

Link	Criterion	PC	TN	Uniform
GL	DIC	65.811	65.605	65.662
	EAIC	32.919	32.797	32.852
	EBIC	33.157	33.035	33.090
	WAIC	29.773	29.274	29.535
PN	DIC	64.515	64.712	64.789
	EAIC	32.360	32.439	32.489
	EBIC	32.598	32.678	32.727
	WAIC	28.543	28.717	28.815

Most of the criteria indicate that the PN link would be more adequate in this case. Additionally, considering the different priors, it can be seen that these measures are very similar in the different models fitted. This is expected, since we fitted the same model under different priors and their final impact on the model fit is subtle.

Additionally, as reference, we also fit the traditional Logistic (logit) and Normal (probit) links considering a quadratic effect and intercept. The respective versions without intercept were also fitted, but did not converge. We found that the restriction that the linear predictor must pass to the origin when no intercept is included in the model combined with a symmetric link was too prohibitive and the model was not able to fit these data well.

Table 11 – Bladder cancer data. Information criteria for the logit (L) and normal (N) links considering an intercept for the linear predictor.

Link	DIC	EAIC	EBIC	WAIC
LI	66.430	33.146	33.384	29.603
NI	64.673	32.466	32.704	28.670

It can be seen from the Table 11 that the probit link with an intercept term presents results closer to the models discussed above. Nonetheless, there is still a difference when compared with the PN link, so we selected this model as the best for the data.

Additionally, Table 12 displays the bladder cancer rates fitted according to the models studied. The results show that the LI and NI models greatly underestimated the cancer rates of the control and lowest-dose groups. Also, the GL model under different priors overestimated the cancer rates of the control and lowest-dose groups. In order to better predict the cancer rate in the unexposed control group, we compute $\sqrt{MSE}$ (see, Subsection 1.1.9). When considering this criterion, the best model was the PN model with PC prior.

Table 12 – Bladder cancer data. Observed and fitted bladder cancer rates $\times 10^5$ for mice exposed to 2-AAF continuously. Models were fitted using the full dataset in Table 9.

	Dose								
Model	0	30	35	45	60	75	100	150	$\sqrt{MSE}$
LI	510.5	820.7	994.3	1604.5	4128.5	13228.0	70605.0	99985.4	1521.30
NI	911.5	733.0	875.1	1463.6	4286.0	14536.3	68870.8	100000.0	876.84
GL_{PC}	1669.4	612.3	701.0	1174.0	3992.5	15622.8	66785.2	99789.0	692.53
GL_{TN}	1031.1	591.1	713.6	1283.0	4419.2	16208.5	64546.1	99624.5	1301.94
GL_U	1372.9	589.7	689.5	1190.8	4104.4	15787.5	65992.5	99740.2	851.69
PN_{PC}	1389.1	662.8	767.0	1268.5	3981.1	14794.6	67847.7	99997.8	687.93
PN_{TN}	998.4	614.3	734.4	1274.9	4114.7	14889.3	67030.4	99995.4	721.50
PN_U	1104.5	626.5	741.3	1267.4	4054.8	14739.6	67165.4	99996.2	728.02
Group rate	990.1	1128.7	0	1941.7	3030.3	16000	67741.9	100000	

Fitted models using all data: x is dose in $^{ppm}/_{100}$:

LI: Logistic with intercept: $\left\{1+\exp\left(-5.27-0.36x+6.51x^2\right)\right\}^{-1}$
NI: Probit with intercept: $\Phi\left(-2.36-1.60x+4.45x^2\right)$
GL_{PC}: Generalized Logistic (PC prior): $\left\{1+\exp\left(--2.63x+5.28x^2\right)\right\}^{5.90}$
GL_{TN}: Generalized Logistic (Normal prior): $\left\{1+\exp\left(-1.92x+4.59x^2\right)\right\}^{6.59}$
GL_U: Generalized Logistic (Uniform prior): $\left\{1+\exp\left(-2.36x+5.03x^2\right)\right\}^{6.18}$
PN_{PC}: Power Probit (PC prior): $\Phi\left(-1.34x+2.89x^2\right)^{6.17}$
PN_{TN}: Power Probit (Normal): $\Phi\left(-1.09x+2.66x^2\right)^{6.65}$
PN_U: Power Probit (Uniform): $\Phi\left(-1.17x+2.72x^2\right)^{6.50}$

In order to compare with more details the performance of the priors studied in the PN model, we considered a predictive modeling approach. In Table 13 we show different prediction measures commonly used in binary regression, some of them, such as as the Critical Success Index (CSI), recommended when the response variable is unbalanced (de la Cruz *et al.*, 2019). The binomial data were treated as having a Bernoulli distribution to get these measures, and except for the $PDIF$, the model with the largest values of these criteria should be chosen since this indicates the strongest similarity between the observed and predicted responses. More detailed information about these measures is available in Subsection 1.1.9.

Table 13 – Bladder cancer data. Predictive measures for the PN link with the PC, Gaussian and uniform priors.

Prior	Value		Predicted 0	Predicted 1	AUC	ACC	TPR	TNR	CSI	SSI	FAITH	PDIF (10^{-3})
PC prior	Observed	0	903	73	0.9	0.92	0.82	0.93	0.35	0.21	0.48	2.75
		1	10	44								
TN	Observed	0	839	137	0.9	0.86	0.85	0.86	0.24	0.14	0.45	4.13
		1	8	46								
Uniform	Observed	0	839	137	0.9	0.86	0.85	0.86	0.24	0.14	0.45	4.13
		1	8	46								

Notes: ACC: accuracy, TPR: sensitivity, TNR: specificity, AUC: area under the curve, CSI: critical success index, SSI: Sokal & Sneath index, FAITH: Faith index and PDIF: pattern difference

From the obtained results, the fitted model using the PN link and the PC prior presented better values of these criteria compared to the Gaussian and uniform priors confirming results discussed previously. This is verified by the ACC, TNR, CSI, SSI, FAITH and PDIF criteria. The PN with the PC prior does not present the better performance in the TPR metric. Again, the results confirmed that the PN link with the PC prior is the best model to predict bladder cancer rates.

On the other hand, for the selected binomial regression model with PN link, Table 14 shows the Bayesian estimates and the 95% credible intervals (in parentheses) for each parameter considering the different priors for the parameter α.

Table 14 – Bladder cancer data. Mean posterior estimation and 95% credible interval in the parentheses for the PN link using the considered priors.

	Prior		
Estimate	PC	TN	Uniform
β_1	-1.339 (-2.673, -0.064)	-1.093 (-2.346, 0.227)	-1.167 (-2.559, 0.117)
β_2	2.886 (1.548, 4.170)	2.661 (1.406, 3.983)	2.728 (1.409, 4.089)
α	6.170 (4.236, 8.421)	6.646 (4.443, 9.037)	6.500 (4.359, 8.959)

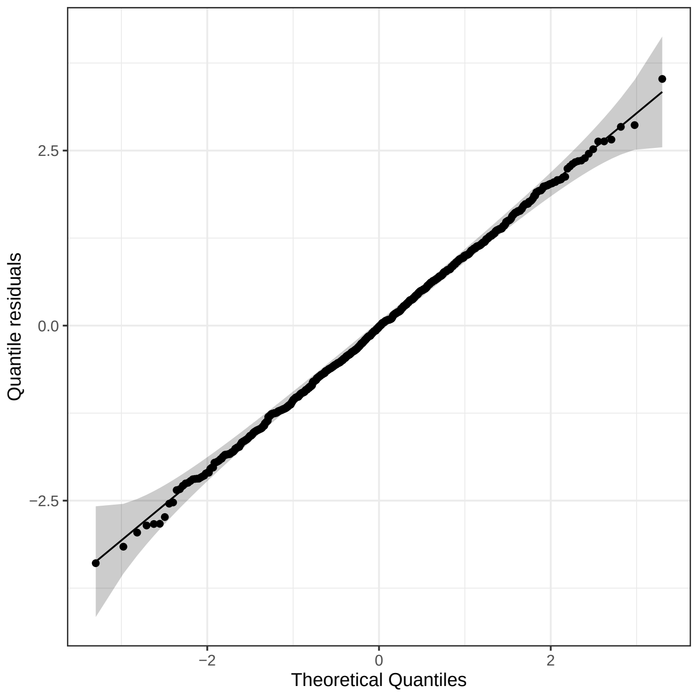
The results for each prior studied confirmed that an asymmetric link is adequate for these data because the credibility intervals obtained for α do not contain the baseline model ($\alpha = 1$). We highlight that shorter intervals for α were obtained using the PC prior in comparison to the others. Another important observation is that unlike the Gaussian and uniform priors, the PC prior has a significant linear term β_1 in comparison with the other priors. This is remarkable from the standpoint of interpreting the results, since considering only a quadratic trend without a linear effect might affect the conclusions if its absence is not justified.

Finally, we include a short diagnostic of the selected model considering the normalized randomized quantile residuals proposed by Dunn & Smyth (1996) for the binary response. Figure 6 shows the simulated envelope of residuals for the selected model considering the PN link under the PC prior. There are no observations lying outside the envelope. This provides further evidence of the adequacy of the PN link using the PC prior. Figure 20 (Appendix C2) shows the scatter plot of the residuals and their distribution. This figure also support the suitability of the model fitted using the PN link and the PC prior. A brief description of this residual method can be found in Subsection 1.1.10.

5.3 Discussion

In this chapter, general and specific expressions of the PC prior were derived for the power link family. This family of links functions is widely applied to binary, binomial and bounded regressions. Moreover, as discussed in Section 4.2.1, we believe this family of priors is constructed based on meaningful principles and in our case offers a straightforward strategy to select the complexity measure λ when $d(\alpha)$ is used as an interpretable transformation of this hyperparameter. This observation is advantageous since enables elucidating prior knowledge about the skewness parameter. Among the results established in this work, we showed the non-identifiability of the power complementary log-log and its associated reversal link. This was done by using the moments generating function of the power Gumbel distribution jointly with the parameterization proposed by van Niekerk & Rue (2020).

Simulation studies showed that, in terms of relative error, length of credibility intervals and coverage probabilities, the PC prior presented better frequentist properties across the considered priors, especially for smaller sample sizes. From the real data application, the PN link using the PC prior with a quadratic trend was the best according to most of the selection criteria used. A comparison with the traditional logit and probit links with an intercept in the linear predictor was also studied, and according to the model selection criteria the PN link with the PC prior was still preferable. Different prediction measures were also used and indicated that the PN link with the PC prior provided then best prediction capacity among the compared models and priors. Also, residual analysis corroborated the appropriate fit of the selected model. Another interesting result is that out of the priors used in the fitting procedure, the PC prior was the only one able to detect the linear term in the quadratic predictor significantly. And, as discussed in Section 5.2, this is preferable in terms of model interpretation.



(a)

Figure 6 – Bladder cancer data: Envelope considering quantile residuals for the PN link under the PC prior

6 Concluding remarks

In this thesis, two default priors were proposed to perform Bayesian estimation considering two different kinds of responses.

With respect to the Student-t spatial regression model, a reference based and two Jeffreys priors were described, establishing conditions which allow them to generate proper posterior distributions. The applicability of the introduced priors was assessed through simulation studies and a real dataset on calcium content in the soil. Limitations in the number of degrees of freedom and the absence of a nugget effect of this method were also mentioned. We hope solutions will be found in future research. An `R` package `OBASpatial` was also developed and is available on `CRAN` for download.

As other research topics, different generalizations of the family of correlation functions are of clear interest, e.g, ansiotropic functions and/or separable ones. Another promising avenue for future research is to pursue a more flexible class of distributions, such as the multivariate skew-t distribution (Azzalini & Capitanio, 2003).

Regarding the binary regression case, weakly informative priors were introduced for the asymmetry parameter of the family of power links. This procedure relied on the univariate family of PC priors developed by Simpson *et al.* (2017). Particular cases, such as the power logit and power probit were studied, obtaining straightforward expressions for the PC prior in these cases. Additionally, non-identifiability of the power Gumbel and its reversal link was shown through the use of the moment generating function and the parameterization proposed by van Niekerk & Rue (2020). As in the Student-t case, a simulation study and a real data application involving bladder cancer in mice were used to assess the performance of the proposed procedure.

As further studies, the inclusion of random effects in the linear predictor might be taken into account when modeling since in many phenomena, binary or bounded responses can be indexed or correlated, e.g., by temporal or spatial patterns. Additionally, the observed identifiability issue between the intercept and skewness parameter for the power links should be better explored. Finally, the development of an `R` package including all the routines regarding the PC priors for this family of link functions is also of interest, due to the interesting properties mentioned.

Bibliography

Aldrich, J. (2008). R. A. Fisher on Bayes and Bayes' theorem. *Bayesian Analysis*, **3**(1), 161 – 170. Cited in page 14.

Ando, T. (2011). Predictive Bayesian model selection. *American Journal of Mathematical and Management Sciences*, **31**(1-2), 13–38. Cited in page 23.

Astrahan, M. (2008). Some implications of linear-quadratic-linear radiation dose-response with regard to hypofractionation. *Med. Phys*, **35**(9), 4161–4172. Cited in page 69.

Atkinson, A. (1985). *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Oxford: Clarendon Press. Cited in page 25.

Azzalini, A. & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society, Series B*, **65**(2), 367–389. Cited 2 times in pages 26 and 74.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC. Cited 2 times in pages 17 and 40.

Bayes, C. L., Bazán, J. L. & de Castro, M. (2017). A quantile parametric mixed regression model for bounded response variables. *Statistics and its Interface*, **19**, 483–493. Cited in page 53.

Bazán, J. L., Bolfarine, H. & Branco, M. D. (2010). A framework for skew-probit links in binary regression. *Communications in Statistics - Theory and Methods*, **39**(4), 678–697. Cited in page 53.

Bazán, J. L., Torres, F., Suzuki, A. & Louzada, F. (2017). Power and reversal power links for binary regressions: an application for motor insurance policyholders. *Applied Stochastic Models in Business and Industry*, **33**(1), 22–34. Cited 3 times in pages 23, 53, and 63.

Berger, J., O. & Bernardo J., M. (1991). On the development of reference priors. *Bayesian Statistics*, **4**, 35–60. Cited 3 times in pages 17, 26, and 30.

Berger, J., O., De Oliveira, V. & Sansó, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96**(456), 1361–1374. Cited 11 times in pages 13, 14, 26, 28, 32, 34, 38, 40, 49, 51, and 82.

Bernardo, J. M. (1979). Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, **41**(2), 113–128. Cited 2 times in pages 14 and 17.

Branco, M. D., Genton, M. G. & Liseo, B. (2013). Objective Bayesian analysis of skew-t distributions. *Scandinavian Journal of Statistics*, **40**(1), 63–85. Cited 3 times in pages 26, 28, and 42.

Brooks, S. P. (2002). Discussion on the paper by spiegelhalter, best, carlin and van der linde. *Journal of The Royal Statistical Society Series B-statistical Methodology*, **64**, 616–639. Cited in page 23.

Chen, M.-H., Dey, D. K. & Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, **94**(448), 1172–1186. Cited in page 52.

Chen, M.-H., Dey, D. K. & Shao, Q.-M. (2001). Bayesian analysis of binary data using skewed logit models. *Calcutta Statistical Association Bulletin*, **51**(1-2), 11–30. Cited in page 53.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, Revised Edition*. John Wiley & Sons, Inc. ISBN 9780471002550. Cited in page 26.

de la Cruz, A., Bazán, J. L., Cancho, V. G. & Dey, D. K. (2019). Performance of asymmetric links and correction methods for imbalanced data in binary regression. *Journal of Statistical Computation and Simulation*, **89**(9), 1694–1714. Cited 3 times in pages 24, 52, and 71.

De Oliveira, V. (2007). Objective Bayesian analysis of spatial data with measurement error. *Canadian Journal of Statistics*, **35**(2), 283–301. Cited 3 times in pages 26, 28, and 42.

Devidas, M., George, E. O. & Zelterman, D. (1993). Generalized logistic models for low—dose response data. *Statistics in Medicine*, **12**(9), 881–892. Cited in page 69.

Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B*, **195**(2), 216–222. Cited in page 20.

Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244. Cited 2 times in pages 25 and 72.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, **27**(8), 861–874. Cited in page 24.

Ferrari, S. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**(7), 799–815. Cited in page 53.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80**(1), 27–38. Cited in page 52.

Flores, S. E., Prates, M. O., Bazán, J. L. & Bolfarine, H. B. (2021). Spatial regression models for bounded response variables with evaluation of the degree of dependence. *Statistics and its Interface*, **14**, 95–107. Cited in page 53.

Fonseca, T. C., R. Ferreira, M. A. & Migon, H. S. (2008). Objective Bayesian analysis for the Student-t regression model. *Biometrika*, **101**(1), 252–252. Cited 2 times in pages 26 and 28.

Galvis, D. M., Bandyopadhyay, D. & Lachos, V. H. (2014). Augmented mixed beta regression models for periodontal proportion data. *Statistics in Medicine*, **33**(21), 3759–3771. Cited in page 53.

Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**(4), 457–472. Cited in page 69.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741. Cited 2 times in pages 18 and 20.

Geweke, J. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, **8**(S1), S19–S40. Cited in page 28.

Gilks, W. R. & Wild, P. (1992). Adaptive rejection sampling for gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **41**(2), 337–348. Cited in page 17.

Gradshtejn, I. S. & Ryzhik, I. M. (1965). *Table of integrals, series and products*. Academic Press. Cited in page 16.

Gupta, R. & Kundu, D. (2010). Generalized logistic distribution. *Journal of Applied Statistical Science*, **18**(1), 51–66. Cited in page 58.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109. Cited 2 times in pages 17 and 18.

He, D., Sun, D. & He, L. (2020). Objective Bayesian analysis for the Student-t linear regression. *Bayesian Analysis*. Cited 3 times in pages 27, 28, and 42.

Ho, H. J., Lin, T.-I., Chen, H.-Y. & Wang, W.-L. (2012). Some results on the truncated multivariate t distribution. *Journal of Statistical Planning and Inference*, **142**(1), 25–40. Cited in page 50.

Hoffman, M. D. & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, **15**(1), 1593–1623. Cited in page 21.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **186**(1007), 453–461. Cited in page 14.

Kaufman, C. G. & Shaby, B. A. (2013). The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, **100**(2), 473–484. Cited in page 45.

Kazianka, H. & Pilz, J. (2012). Objective Bayesian analysis of spatial data with uncertain nugget and range parameters. *Canadian Journal of Statistics*, **40**(2), 304–327. Cited 4 times in pages 26, 28, 42, and 51.

Kennedy, M. C. & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B*, **63**(3), 425–464. Cited in page 26.

Kotz, S. & Nadarajah, S. (2004). *Multivariate T-Distributions and Their Applications*. Cambridge University Press. Cited in page 40.

Landau, E. (1909). *Handbuch der Lehre von der Verteilung der Primzahlen*. Teubner. Cited in page 21.

Lange, K. L., Little, R. J. & Taylor, J. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, **84**(408), 881–896. Cited 2 times in pages 34 and 35.

Lee, D. H. & Sinha, S. (2019). Identifiability and bias reduction in the skew-probit model for a binary response. *Journal of Statistical Computation and Simulation*, **89**, 1621–1648. Cited 2 times in pages 52 and 55.

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in bayesian analyses. *Oikos*, **128**(7), 912–928. Cited in page 14.

Lemonte, Artur J.; Bazán, J. L. (2017). New links for binary regression: an application to coca cultivation in peru. *TEST*. Cited 4 times in pages 23, 25, 52, and 53.

Leonard, T. H. (2014). A personal history of bayesian statistics. *WIREs Computational Statistics*, **6**(2), 80–115. Cited in page 14.

Lobo, V. G. & Fonseca, T. C. (2020). Bayesian residual analysis for spatially correlated data. *Statistical Modelling*, **20**(2), 171–194. Cited in page 40.

McCullagh, P. & Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Chapman & Hall. Cited in page 22.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, M. & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**(6), 1087–1092. Cited 3 times in pages 17, 18, and 20.

Muré, J. (2020). Propriety of the reference posterior distribution in gaussian process modeling. Cited in page 34.

Naranjo, L., Martín, J. & Pérez, C. (2014). Bayesian binary regression with exponential power link. *Computational Statistics and Data Analysis*, **71**, 464–476. Cited in page 53.

Neal, R. (2011). *Handbook of Markov Chain Monte Carlo*, chapter 5: MCMC Using Hamiltonian Dynamics. CRC Press. Cited in page 20.

Oliveira, M. C. N. (2003). *Métodos de estimação de parâmetros em modelos geoestatísticos com diferentes estruturas de covariâncias: uma aplicação ao teor de cálcio no solo*. Ph.D. thesis, ESALQ/USP/Brasil. Cited in page 47.

Ordoñez, A., Prates, M. O., Matos, L. A. & Lachos, V. H. (2021). *OBASpatial: Objective Bayesian Analysis for Spatial Regression Models*. R package version 1.8. Cited in page 40.

Palacios, M. B. & Steel, M. F. J. (2006). Non-gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, **101**(474), 604–618. Cited 2 times in pages 15 and 27.

Paulo, R. (2005). Default priors for gaussian processes. *The Annals of Statistics*, **33**(2), 556–582. Cited in page 26.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Cited in page 64.

Ren, C., Sun, D. & He, C. (2012). Objective Bayesian analysis for a spatial model with nugget effects. *Journal of Statistical Planning and Inference*, **142**(7), 1933–1946. Cited 3 times in pages 26, 28, and 51.

Ren, C., Sun, D. & Sahu, S. K. (2013). Objective Bayesian analysis of spatial models with separable correlation functions. *Canadian Journal of Statistics*, **41**(3), 488–507. Cited 2 times in pages 26 and 28.

Ribeiro Jr, P. J., Diggle, P. J., Schlather, M., Bivand, R. & Ripley, B. (2020). *geoR: Analysis of Geostatistical Data*. R package version 1.8-1. Cited in page 47.

Robert, C. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. Springer, New York. Cited in page 14.

Robert, C. P. & Rousseau, J. (2017). How principled and practical are penalised complexity priors? **32**(1), 36–40. Cited in page 55.

Simpson, D., Rue, H., Riebler, A., Martins, T. G. & Sørbye, S. H. (2017). Penalising model component complexity: a principled, practical approach to constructing priors. *Statistical Science*, **32**(1), 1–28. Cited 6 times in pages 14, 15, 52, 53, 54, and 74.

Stan Development Team (2020). The Stan Core Library. Version 2.26.0. Cited in page 64.

Stein, C. (1985). On the coverage probability of confidence sets based on a prior distribution. *Sequential Methods in Statistics.*, pages 485–514. Cited in page 42.

Syversveen, A. R. (1998). Noninformative bayesian priors. interpretation and problems with construction and applications. Cited in page 14.

van Niekerk, J. & Rue, H. (2020). Skewed probit regression – identifiability, contraction and reformulation. Cited 4 times in pages 55, 63, 73, and 74.

Villa, C. & Walker, S. G. (2014). Objective prior for the number of degrees of freedom of a t distribution. *Bayesian Analysis*, **9**(1), 197–220. Cited in page 27.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.*, **11**, 3571–3594. Cited in page 23.

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**(354), 400–405. Cited 2 times in pages 26 and 40.

APPENDIX A – Appendix for chapter 2: proofs of propositions and lemmas.

A1 Technical details

First, we state some useful properties that will be needed to proof the main results of the manuscript.

1. Consider $\mathbf{Y} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ and $\mathbf{A}, \mathbf{B}$ two symmetric matrices. Then,
   a) $E\left\{\mathbf{Y}^\top \mathbf{A} \mathbf{Y}\right\} = \boldsymbol{\mu}^\top \mathbf{A} \boldsymbol{\mu} + \frac{\nu}{\nu - 2} tr\left[\mathbf{A}\boldsymbol{\Sigma}\right],$
   b)
$$\begin{aligned} Cov\left\{\mathbf{Y}^\top \mathbf{A}\mathbf{Y}, \mathbf{Y}^\top \mathbf{B}\mathbf{Y}\right\} &= \frac{2\nu^2}{(\nu-2)(\nu-4)} tr\left[\mathbf{A}\boldsymbol{\Sigma}\mathbf{B}\boldsymbol{\Sigma}\right] \\ &+ \frac{2\nu^2}{(\nu-2)^2(\nu-4)} tr\left[\mathbf{A}\boldsymbol{\Sigma}\right] tr\left[\mathbf{B}\boldsymbol{\Sigma}\right]. \end{aligned}$$
2. Suppose that $X \sim F(d_1, d_2)$. Then, X has density given by
$$f(X) = \frac{\Gamma(d_1 + d_2)}{\Gamma(d_1)\Gamma(d_2)} \left(\frac{d_1}{d_2}\right)^{\frac{d_1}{2}} X^{\frac{d_1}{2}-1} \left(1 + \frac{d_1}{d_2}X\right)^{-\frac{d_1+d_2}{2}}.$$
3. If $X \sim F(d_1, d_2)$, then $X^{-1} \sim F(d_2, d_1)$.
4. If $X \sim F(d_1, d_2)$, then $\frac{\frac{d_1}{d_2}X}{1 + \frac{d_1}{d_2}X} \sim Beta(\frac{d_1}{2}, \frac{d_2}{2})$.
5. If $X \sim Beta(a, b)$, then $1 - X \sim Beta(b, a)$ (Mirror property).
6. If $X \sim Beta(a, b)$, then
   - $E\left\{X\right\} = \frac{a}{a+b}$
   - $Var\left\{X\right\} = \frac{ab}{(a+b)^2(a+b+1)}$
   - $E\left\{\log(X)\right\} = \Psi(a) - \Psi(a+b)$
   - $E\left\{X\log(X)\right\} = \frac{a}{a+b}\left[\Psi(a+1) - \Psi(a+b+1)\right]$, where $\Psi(.)$ is the digamma function.

A1.1 Conditions of Lemma 2 (Berger *et al.*, 2001)

In this section, we introduce the conditions of Lemma 2 (Berger *et al.*, 2001) necessary for the results presented in the main version of the manuscript. These conditions are the following:

1. The correlation matrix can be decomposed as $\mathbf{R}(\phi) = \mathbf{1}\mathbf{1}^\top + \psi(\phi)\mathbf{D} + \omega(\phi)\mathbf{D}^* + \mathbf{R}^*(\phi)$ and $(\partial/\partial\phi)\,\mathbf{R}(\phi) = \psi'(\phi)\mathbf{D} + \omega(\phi)'\mathbf{D}^* + (\partial/\partial\phi)\,\mathbf{R}^*(\phi)$, with $\mathbf{1} = (1,\ldots,1)^\top$, $\mathbf{D}$ (nonsigular) and $\mathbf{D}^*$ fixed matrices; and $\psi(\phi)\,(>0)$, $\omega(\phi)$ and $\mathbf{R}^*(\phi)$ differentiable.

2. As $\phi \to \infty$,

$$\frac{\omega(\phi)}{\psi(\phi)} \to 0, \quad \frac{\omega'(\phi)}{\psi'(\phi)} \to 0, \quad \frac{\|\mathbf{R}^*(\phi)\|_\infty}{\omega(\phi)}, \to 0, \quad \frac{\|(\partial/\partial\phi)\,\mathbf{R}^*(\phi)\|_\infty}{\omega'(\phi)} \to 0 \quad \text{and}$$

$$\frac{\left\|{}^{(\partial/\partial\phi)\mathbf{R}^*(\phi)_\infty}/_{\psi'(\phi)}\right\| + \left\|{}^{\mathbf{R}^*(\phi)}/_{\psi(\phi)}\right\|_\infty}{{}^{\omega'(\phi)}/_{\psi'(\phi)} - {}^{\omega(\phi)}/_{\psi(\phi)}} \to 0,$$

where $\|A\|_\infty = \max_{i,j} |A_{ij}|$.

The presented hypotheses are satisfied by following families of correlation functions: the spherical family, $\psi(\phi) = \phi^{-1}$ and $\omega(\phi) = \phi^{-3}$; the power exponential family, $\psi(\phi) = \phi^{-\kappa}$ and $\omega(\phi) = \phi^{-2\kappa}$; the Cauchy family, $\psi(\phi) = \phi^{-2}$ and $\omega(\phi) = \phi^{-4}$; and the Matérn family where the expressions for $\psi(\phi)$ and $\omega(\phi)$ are given in Appendix C of Berger *et al.* (2001).

A2 Behavior of $A(\nu)$

In this section we analyze the behaviour of $A(\nu)$, considering that the Student-t distribution converges to the normal distribution as $\nu \to +\infty$. By Stirling formula, it follows that

$$\begin{aligned}
\Gamma(\frac{\nu}{2} - 1 + 1) &\sim \sqrt{2\pi\left(\frac{\nu}{2} - 1\right)}\left(\frac{\frac{\nu}{2} - 1}{e}\right)^{\frac{\nu}{2}-1}, \\
\Gamma(\frac{\nu}{2} - a + 1) &\sim \sqrt{2\pi\left(\frac{\nu}{2} - a\right)}\left(\frac{\frac{\nu}{2} - a}{e}\right)^{\frac{\nu}{2}-a}.
\end{aligned}$$

Then,

$$\begin{aligned}
\lim_{\nu\to+\infty} A(\nu) &= \lim_{\nu\to+\infty} \nu^{-(1-a)} \frac{\Gamma(\frac{\nu}{2} - a + 1)}{\Gamma(\frac{\nu}{2})} \\
&= \lim_{\nu\to+\infty} \nu^{-(1-a)} \frac{\sqrt{2\pi\left(\frac{\nu}{2} - a\right)}\left(\frac{\frac{\nu}{2}-a}{e}\right)^{\frac{\nu}{2}-a}}{\sqrt{2\pi\left(\frac{\nu}{2} - 1\right)}\left(\frac{\frac{\nu}{2}-1}{e}\right)^{\frac{\nu}{2}-1}} \\
&= \lim_{\nu\to+\infty} \frac{\left(\frac{\nu}{2} - a\right)^{-1}}{\left(\frac{\nu}{2} - 1\right)^{-1}} \sqrt{\frac{2\pi\left(\frac{\nu}{2} - a\right)}{2\pi\left(\frac{\nu}{2} - 1\right)}}\, \nu^{-(1-a)} \left(\frac{\frac{\nu}{2} - a}{\frac{\nu}{2} - 1}\right)^{\frac{\nu}{2}} \left(\frac{\nu}{2} - a\right)^{1-a}
\end{aligned}$$

$$= \lim_{\nu\to+\infty} \nu^{-(1-a)} \left(\frac{\frac{\nu}{2}-a}{\frac{\nu}{2}-1}\right)^{\frac{\nu}{2}} \left(\frac{\nu}{2}-a\right)^{1-a}.$$

The last equality follows because the first two terms of the limit tends to 1, as $\nu \to +\infty$. We have also that $\left(\frac{\frac{\nu}{2}-a}{\frac{\nu}{2}-1}\right)^{\frac{\nu}{2}} \to 1$ as $\nu \to +\infty$ (it can be easily showed by studying the limit of its natural logarithm and using L'Hôpital's Rule). Then,

$$\lim_{\nu\to+\infty} A(\nu) = \lim_{\nu\to+\infty} \nu^{-(1-a)} \left(\frac{\nu}{2}-a\right)^{1-a} = \lim_{\nu\to+\infty} \frac{1}{2^{1-a}} \left(\frac{\nu-2a}{\nu}\right)^{1-a} = \frac{1}{2^{1-a}}.$$

Therefore, the behaviour of $A(\nu)$ is constant, as $\nu \to +\infty$. So, for an appropriate choice of $\pi(\nu, \phi)$, we have that its posterior distribution converges to the normal one.

A3 Proof of Theorem 2.1

Let $\ell_1(\theta^*) = \log(L_1(\theta^*))$ and $S^* = S^2/\sigma^2$, then

$$\begin{aligned}\ell_1(\theta^*) &= \log\left\{\Gamma\left(\frac{\nu+n_p}{2}\right)\right\} - \log\left\{\Gamma\left(\frac{\nu}{2}\right)\right\} - \frac{n_p}{2}\log(\sigma^2) - \frac{1}{2}\log(|\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X}|) \\ &\quad - \left(\frac{\nu+n_p}{2}\right)\log(\nu+n_pS^*) - \frac{1}{2}\log(|\mathbf{R}|) + \frac{\nu}{2}\log(\nu). \qquad \text{(A3.1)}\end{aligned}$$

From Equation (A3.1), we have that

$$\begin{aligned}\frac{\partial\ell_1(\theta^*)}{\partial\sigma^2} &= -\frac{1}{2}\left\{\frac{n_p}{\sigma^2} + \frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\sigma^2}\right\}, \\ \frac{\partial\ell_1(\theta^*)}{\partial\phi} &= -\frac{1}{2}\left\{\frac{\nu+n_p}{\nu+n_pS^*}n_p\frac{\partial S^*}{\partial\phi} - tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{X}\right]\right. \\ &\quad \left. + \, tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\right]\right\}, \\ \frac{\partial\ell_1(\theta^*)}{\partial\nu} &= \frac{1}{2}C(\nu) + \frac{1}{2}\log\left(\frac{1}{\nu+n_pS^*}\right) - \frac{n_p+\nu}{2(\nu+n_pS^*)} + \frac{1}{2}\log(\nu) + \frac{1}{2}.\end{aligned}$$

Define $I_1(\theta^*)$, as

$$I_1(\theta^*) = E\left\{\left(\frac{\partial\ell_1(\theta^*)}{\partial\theta^*}\right)\left(\frac{\partial\ell_1(\theta^*)}{\partial\theta^*}\right)^\top\right\}.$$

Now, we proceed to calculate each component of the matrix $I_1(\theta^*)$. Thus, we have

$$\begin{aligned}E\left\{\left(\frac{\partial\ell_1(\theta^*)}{\partial\sigma^2}\right)^2\right\} &= \frac{1}{4\sigma^4}\left[n_p^2 - 2n_p^2(\nu+n_p)E\left\{\frac{S^*}{\nu+n_pS^*}\right\} + n_p^2(\nu+n_p)^2\times\right. \\ &\quad \left. E\left\{\left(\frac{S^*}{\nu+n_pS^*}\right)^2\right\}\right]. \qquad \text{(A3.2)}\end{aligned}$$

It can be shown that $\frac{\sigma^2}{S^2} \sim F(\nu, n_p)$, then $S^* = \frac{S^2}{\sigma^2} \sim F(n_p, \nu)$. Now, it follows that

$$E\left\{\frac{S^*}{\nu+n_pS^*}\right\} = E\left\{\frac{\frac{\nu}{n_p}\frac{n_p}{\nu}S^*}{\nu\left(1+\frac{n_p}{\nu}S^*\right)}\right\} = \frac{1}{n_p}E\left\{\frac{\frac{n_p}{\nu}S^*}{1+\frac{n_p}{\nu}S^*}\right\} = \frac{1}{n_p}E\left\{1 - \frac{1}{1+\frac{n_p}{\nu}S^*}\right\}$$

$$= \quad E\left\{1-Q\right\},$$

where $Q \sim Beta(\frac{\nu}{2}, \frac{n_p}{2})$. Then, $E\left\{\frac{S^*}{\nu + n_p S^*}\right\} = \frac{n_p}{n_p+\nu}\frac{1}{n_p} = \frac{1}{n_p+\nu}$. Moreover,

$$E\left\{\left(\frac{S^*}{\nu + n_p S^*}\right)^2\right\} = \frac{1}{n_p^2}E\left\{\left(\frac{\frac{n_p}{\nu}S^*}{1+\frac{n_p}{\nu}S^*}\right)^2\right\} = \frac{1}{n_p^2}E\left\{(1-Q)^2\right\}.$$

In this case, we have by the mirror property that $1-Q \sim Beta(\frac{n_p}{2}, \frac{\nu}{2})$, thus, $E\left\{1-Q\right\} = \frac{n_p}{n_p+\nu}$, $V\left\{1-Q\right\} = \frac{2\nu n_p}{(n_p+nu_0)^2(\nu+n_p+1)}$. Then,

$$\begin{aligned}
E\left\{\left(\frac{S^*}{\nu + n_p S^*}\right)^2\right\} &= \frac{1}{n_p^2}\left\{\frac{2\nu n_p}{(n_p+\nu)^2(\nu+n_p+1)} + \frac{n_p^2}{(n_p+\nu)^2}\right\} \\
&= \frac{1}{n_p^2(\nu+n_p)^2}\left\{\frac{2n_p\nu}{\nu+n_p+2} + n_p^2\right\},
\end{aligned}$$

and replacing each expectation in Equation(A3.2), we have that

$$E\left\{\left(\frac{\partial \ell_1(\theta^*)}{\partial\sigma^2}\right)^2\right\} = \frac{1}{2\sigma^4}\frac{\nu n_p}{\nu+n_p+2} = \frac{1}{2\sigma^4}B. \tag{A3.3}$$

Since $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{R}^{-1}X)^{-1}\mathbf{X}^\top\mathbf{R}^{-1}$ and $\mathbf{Z}_1 = {}^{(\mathbf{Y}-\mathbf{X}\hat{\beta})}/_{\sigma} \sim t_n(\mathbf{0}, \mathbf{PR}, \nu)$, then

$$\begin{aligned}
\frac{\partial \ell_1(\theta^*)}{\partial\phi}\frac{\partial \ell_1(\theta^*)}{\partial\sigma^2} &= \frac{1}{4\sigma^2}\Bigg\{-\frac{(\nu+n_p)n_p}{(\nu+n_pS^*)}\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{Z}_1 + n_p tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\right] \\
&- n_p tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{X}\right] \\
&+ \frac{(\nu+n_p)^2 n_p}{(\nu+n_pS^*)^2}S^*\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{Z}_1 \\
&+ \frac{(\nu+n_p)n_pS^*}{(\nu+n_pS^*)}tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{X}\right] \\
&- \frac{(\nu+n_p)n_pS^*}{(\nu+n_pS^*)}tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\right]\Bigg\}
\end{aligned}$$

and

$$\begin{aligned}
E\left\{\frac{\partial \ell_1(\theta^*)}{\partial\phi}\frac{\partial \ell_1(\theta^*)}{\partial\sigma^2}\right\} &= \frac{1}{4\sigma^2}\Bigg\{-E\left\{\frac{(\nu+n_p)n_p}{(\nu+n_pS^*)}\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} \\
&- n_p tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{X}\right] \\
&+ E\left\{\frac{(\nu+n_p)^2 n_p}{(\nu+n_pS^*)^2}S^*\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} \\
&+ tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{X}\right]E\left\{\frac{(\nu+n_p)n_pS^*}{(\nu+n_pS^*)}\right\} \\
&+ n_p tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\right]
\end{aligned} \tag{A3.4}$$

$$- \quad tr\left[\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\right] E\left\{\frac{(\nu+n_p)n_p S^*}{(\nu+n_p S^*)}\right\}\Bigg\}. \tag{A3.5}$$

Now, we proceed to compute each expectation value involved in Equation (A3.4). Note first that,

$$E\left\{\frac{(\nu+n_p)n_p}{(\nu+n_p S^*)}\mathbf{Z}_1^\top \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} = E\left\{\frac{(\nu+n_p)n_p}{(\nu+n_p S^*)}E\left\{\mathbf{Z}_1^\top \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1|S^*\right\}\right\}.$$

Hence, by using the Property 1 of Section A1, it follows that

$$\begin{aligned}
E\left\{\frac{(\nu+n_p)n_p}{(\nu+n_p S^*)}\mathbf{Z}_1^\top \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} &\approx \frac{\nu}{\nu-2}E\left\{\frac{(\nu+n_p)n_p}{(\nu+n_p S^*)}tr\left[\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\mathbf{R}\right]\right\} \\
&\approx \frac{\nu n_p(\nu+n_p)}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right]E\left\{\frac{1}{\nu+n_p S^*}\right\} \\
&\approx \frac{n_p(\nu+n_p)}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial \phi}R^{-1}\mathbf{P}\right]E\left\{\frac{1}{1+\frac{n_p}{\nu}S^*}\right\} \\
&\approx \frac{n_p(\nu+n_p)}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right]E\left\{Q\right\},
\end{aligned}$$

where $Q \sim Beta(\frac{\nu}{2}, \frac{n_p}{2})$. Then,

$$E\left\{\frac{(\nu+n_p)n_p}{(\nu+n_p S^*)}\mathbf{Z}_1^\top \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} \approx \frac{\nu n_p}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right].$$

Second, we have that

$$\begin{aligned}
E\left\{\frac{(\nu+n_p)^2 n_p}{(\nu+n_p S^*)^2}S^*\mathbf{Z}_1^\top \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} &= E\left\{\frac{(\nu+n_p)^2\nu}{(\nu+n_p S^*)^2}S^*\right. \\
&\quad \left. E\left\{\mathbf{Z}_1^\top \mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1|S^*\right\}\right\} \\
&\approx \frac{\nu n_p(\nu+n_p)^2}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right] \\
&\quad E\left\{\frac{S^*}{\nu+n_p S^*}\frac{1}{\nu+n_p S^*}\right\}.
\end{aligned}$$

Moreover,

$$E\left\{\frac{S^*}{\nu+n_p S^*}\frac{1}{\nu+n_p S^*}\right\} = \frac{1}{n_p\nu}E\left\{Q_c(1-Q_c)\right\},$$

where $Q_c = \dfrac{\frac{n_p}{\nu}S^*}{1+\frac{n_p}{\nu}S^*} = (1-Q)$ and $Q \sim Beta(\frac{\nu}{2}, \frac{n_p}{2})$. Then, $E\left\{Q_c\right\} = \dfrac{n_p}{n_p+\nu}$ and $E\left\{Q_c^2\right\} = (\dfrac{n_p}{n_p+\nu})^2 + \dfrac{2n_p\nu}{(n_p+\nu)^2(n_p+\nu+2)}$ and therefore,

$$E\left\{\frac{S^*}{(\nu+n_p S^*)^2}\right\} = \frac{1}{n_p\nu}\left(\frac{n_p}{\nu+n_p} - \frac{n_p^2}{(\nu+n_p)^2} - \frac{2n_p\nu}{(n_p+\nu)^2(n_p+\nu+2)}\right)$$

and

$$
\begin{aligned}
E\left\{\frac{(\nu+n_p)^2 n_p}{(\nu+n_pS^*)^2}S^*\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{Z}_1\right\} &\approx \frac{1}{\nu-2}tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right]\times \\
&\left\{(\nu+n_p)n_p - n_p^2 - \frac{2n_p\nu}{(n_p+\nu+2)}\right\} \\
&\approx \frac{n_p\nu}{\nu-2}\left(\frac{n_p+\nu}{n_p+\nu+2}\right)tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right].
\end{aligned}
$$

Third, we have that

$$
E\left\{\frac{(\nu+n_p)n_pS^*}{\nu+n_pS^*}\right\} = \frac{(\nu+n_p)n_p}{\nu}\frac{\nu}{n_p}E\left\{Q_c\right\} = \frac{(\nu+n_p)n_p}{\nu}\frac{\nu}{n_p}\frac{n_p}{\nu+n_p} = n_p.
$$

Replacing the expectations obtained in Equation (A3.4), we have that

$$
\begin{aligned}
E\left\{\frac{\partial\ell_1(\theta^*)}{\partial\phi}\frac{\partial\ell_1(\theta^*)}{\partial\sigma^2}\right\} &\approx \frac{1}{4\sigma^2}\left(\frac{-2\nu n_p}{(n_p+\nu+2)(\nu-2)}tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right]\right) \\
&= \frac{1}{4\sigma^2}B_{11}.
\end{aligned} \tag{A3.6}
$$

Now, we have

$$
\begin{aligned}
E\left\{\frac{\partial\ell_1(\theta^*)}{\partial\sigma^2}\frac{\partial\ell_1(\theta^*)}{\partial\nu}\right\} &= -\frac{n_p}{4\sigma^4}\Bigg[C(\nu) + E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} - E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\} \\
&+ \log(\nu) + 1 - E\left\{\frac{\nu+n_p}{\nu+n_pS^*}S^*\right\}C(\nu) \\
&- E\left\{\frac{\nu+n_p}{\nu+n_pS^*}S^*\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} \\
&+ E\left\{\frac{(n_p+\nu)^2S^*}{(\nu+n_pS^*)^2}\right\} - E\left\{\frac{\nu+n_p}{\nu+n_pS^*}S^*\right\}\log(\nu) \\
&- E\left\{\frac{\nu+n_p}{\nu+n_pS^*}S^*\right\}\Bigg].
\end{aligned} \tag{A3.7}
$$

Since,

$$
E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\} = E\left\{\frac{\nu+n_p}{\nu+n_pS^*}S^*\right\} = 1 \text{ and } E\left\{\frac{(n_p+\nu)^2S^*}{(\nu+n_pS^*)^2}\right\} = \frac{n_p+\nu}{n_p+\nu+2},
$$

it follows that

$$
\begin{aligned}
E\left\{\frac{\nu+n_p}{\nu+n_pS^*}S^*\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} &= \frac{\nu+n_p}{n_p}E\left\{(1-Q)\left(-\log(\nu)+\log(Q)\right)\right\} \\
&= -\log(\nu) - \frac{2}{n_p+\nu} - C(\nu)
\end{aligned}
$$

and

$$
E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} = E\left\{-\log(\nu)+\log(Q)\right\} = -\log(\nu) - C(\nu),
$$

where $Q \sim Beta(\frac{\nu}{2}, \frac{n_p}{2})$. Replacing in (A3.7), we have therefore that

$$E\left\{\frac{\partial \ell_1(\theta^*)}{\partial \sigma^2}\frac{\partial \ell_1(\theta^*)}{\partial \nu}\right\} = \frac{-n_p}{\sigma^2(n_p+\nu)(n_p+\nu+2)} = \frac{1}{\sigma^2}B_{12}. \tag{A3.8}$$

Furthermore,

$$\begin{aligned}
\mathrm{E}\left\{\left(\frac{\partial \ell_1(\theta^*)}{\partial \phi}\right)^2\right\} &= \frac{1}{4}\Bigg\{E\left\{\frac{(\nu+n_p)^2 n_p^2}{(\nu+n_pS^*)^2}\left(\frac{\partial S^*}{\partial \phi}\right)^2\right\} + tr^2\left[\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\right] \\
&+ tr^2\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{X}\right] \\
&- 2E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial \phi}\right\}tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{X}\right] \\
&+ 2E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial \phi}\right\}tr\left[\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\right] \\
&- 2tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{X}\right] \qquad \text{(A3.9)} \\
&\times tr\left[\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\right]\Bigg\}. \qquad \text{(A3.10)}
\end{aligned}$$

Now, we proceed to calculate each expectation involved in Equation (A3.10). First, using the Property 1 of the Section A1 for the Student-t distribution, we have that

$$\begin{aligned}
E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial \phi}\right\} &= \frac{-\nu-n_p}{\nu-2}E\left\{QE\left\{\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\middle|S^*\right\}\right\} \\
&\approx -\frac{\nu}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right].
\end{aligned}$$

Second, we have

$$E\left\{\frac{(\nu+n_p)^2 n_p^2}{(\nu+n_pS^*)^2}\left(\frac{\partial S^*}{\partial \phi}\right)^2\right\} = (\nu+n_p)^2E\left\{\frac{1}{(\nu+n_pS^*)^2}\;E\left\{\left(\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\right)^2|S^*\right\}\right\}$$

and using the Property 1 of Section A1, we have

$$\begin{aligned}
E\left\{\left(\mathbf{Z}_1^\top\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{Z}_1\right)^2\right\} &= \frac{2\nu^2}{(\nu-2)(\nu-4)}tr\left[\left(\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right)^2\right] \\
&+ \frac{2\nu^2}{(\nu-2)^2(\nu-4)}tr^2\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right] + \frac{\nu^2}{(\nu-2)^2}tr^2\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right] \\
&= \frac{\nu^2}{(\nu-2)(\nu-4)}\left\{2tr\left[\left(\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right)^2\right] + tr^2\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right]\right\} \\
&= \mathbf{A}
\end{aligned}$$

and

$$E\left\{\frac{(\nu+n_p)^2 n_p^2}{(\nu+n_pS^*)^2}\left(\frac{\partial S^*}{\partial \phi}\right)^2\right\} \approx (\nu+n_p)^2\mathbf{A}E\left\{\frac{1}{(\nu+n_pS^*)^2}\right\}.$$

Further,

$$V\left\{\frac{1}{\nu+n_pS^*}\right\} = \frac{1}{\nu^2}V\left\{Q\right\} = \frac{2n_p}{(\nu+n_p)^2(\nu+n_p+2)\nu},$$

then

$$\begin{aligned} E\left\{\frac{1}{(\nu+n_pS^*)^2}\right\} = \frac{1}{\nu^2}V\left\{Q\right\} &= \frac{2\nu}{(\nu+n_p)^2(\nu+n_p+2)\nu} + \frac{1}{(n_p+\nu)^2} \\ &= \frac{1}{(n_p+\nu)^2}\left(\frac{2n_p}{(\nu+n_p+2)\nu}+1\right), \end{aligned}$$

so

$$E\left\{\frac{(\nu+n_p)^2n_p^2}{(\nu+n_pS^*)^2}\left(\frac{\partial S^*}{\partial\phi}\right)^2\right\} = \left(\frac{2n_p}{(\nu+n_p+2)\nu}+1\right)\mathbf{A}.$$

Also, we have

$$tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right] = tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\mathbf{R}\right] = tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\mathbf{R}\mathbf{P}^\top\right],$$

and it can be shown that

$$\begin{aligned} tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\mathbf{R}\mathbf{P}^\top\right] &= tr\left[\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\right] \\ &- tr\left[(\mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{X}\right]. \end{aligned} \tag{A3.11}$$

Replacing these expectations along with (A3.11) in Equation (A3.10), we finally have that

$$\begin{aligned} E\left\{\left(\frac{\partial\ell_1(\theta^*)}{\partial\phi}\right)^2\right\} &\approx \frac{1}{4}\left\{\left(\frac{2n_p}{(\nu+n_p+2)\nu}+1\right)\mathbf{A} - \left(\frac{\nu+2}{\nu-2}\right)tr^2\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right]\right\} \\ &= \frac{1}{4}C. \end{aligned} \tag{A3.12}$$

Now,

$$\begin{aligned} E\left\{\frac{\partial\ell_1(\theta^*)}{\partial\nu}\frac{\partial\ell_1(\theta^*)}{\partial\phi}\right\} &= -\frac{1}{4}\Bigg\{C(\nu)E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\phi}\right\} + C(\nu)tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right] \\ &+ E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\phi}\right\} \\ &+ E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\}tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right] \\ &- E\left\{\frac{(n_p+\nu)^2n_p}{(\nu+n_pS^*)^2}\frac{\partial S^*}{\partial\phi}\right\} - E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\}tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right] \\ &+ \log(\nu)E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\phi}\right\} + \log(\nu)tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right] \\ &+ E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\phi}\right\} + tr\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right]\Bigg\}. \end{aligned} \tag{A3.13}$$

Since,

$$\begin{aligned}
E\left\{\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\phi}\right\} &= \frac{\nu}{\nu-2}tr\left[\frac{\partial \mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right],\\
E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} &= -\log(\nu)-C(\nu),\\
E\left\{\frac{(n_p+\nu)^2n_p}{(\nu+n_pS^*)^2}\frac{\partial S^*}{\partial\phi}\right\} &= -\left(\frac{2n_p}{(\nu+n_p+2)(\nu-2)}+\frac{\nu}{\nu-2}\right)\times tr\left[\frac{\partial \mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right],
\end{aligned}$$

and

$$\begin{aligned}
E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\frac{(\nu+n_p)n_p}{\nu+n_pS^*}\frac{\partial S^*}{\partial\phi}\right\} &= -\frac{\nu}{\nu-2}\left(\frac{2n_p}{\nu(\nu+n_p)}-\log(\nu)-C(\nu)\right)\times\\
&\quad tr\left[\frac{\partial \mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right].
\end{aligned}$$

Replacing these expectations in (A3.13), we obtain that

$$E\left\{\frac{\partial\ell_1(\theta^*)}{\partial\nu}\frac{\partial\ell_1(\theta^*)}{\partial\phi}\right\} = \frac{n_p}{(\nu-2)(\nu+n_p)(\nu+n_p+2)}tr\left[\frac{\partial \mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right] = C_{11}. \quad \text{(A3.14)}$$

Lastly,

$$\begin{aligned}
E\left\{\left(\frac{\partial\ell_1(\theta^*)}{\partial\nu}\right)^2\right\} &= \Bigg\{C^2(\nu)+E\left\{\log^2\left(\frac{1}{\nu+n_pS^*}\right)\right\}+E\left\{\frac{(n_p+\nu)^2}{(\nu+n_pS^*)^2}\right\}\\
&+ \log^2(\nu)+1+2C(\nu)E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\}\\
&- 2C(\nu)E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\}+2\log(\nu)C(\nu)\\
&+ 2C(\nu)-2E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\log\left(\frac{1}{\nu+n_pS^*}\right)\right\}\\
&+ 2\log(\nu)E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\}+2E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\}\\
&- 2\log(\nu)E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\}-2E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\}\\
&+ 2\log(\nu)\Bigg\}. \qquad \text{(A3.15)}
\end{aligned}$$

Using the Properties (5) and (6), given in Section A1, we have that

$$\begin{aligned}
E\left\{\log^2\left(\frac{1}{\nu+n_pS^*}\right)\right\} &= (\log(\nu)-C(\nu))^2-C_1(\nu),\\
E\left\{\frac{(n_p+\nu)^2}{(\nu+n_pS^*)^2}\right\} &= 1+\frac{2n_p}{\nu(\nu+n_p+2)},\\
E\left\{\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} &= -\log(\nu)-C(\nu),
\end{aligned}$$

$$\begin{aligned} E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\right\} &= 1, \\ E\left\{\frac{n_p+\nu}{\nu+n_pS^*}\log\left(\frac{1}{\nu+n_pS^*}\right)\right\} &= \frac{2n_p}{\nu(n_p+\nu)}-\log(\nu)-C(\nu). \end{aligned}$$

Replacing these expectations in (A3.15), we have that

$$E\left\{\left(\frac{\partial\ell_1(\theta^*)}{\partial\nu}\right)^2\right\}=\frac{1}{4}\left\{\frac{-2n_p}{\nu}\left(\frac{n_p+\nu+4}{(n_p+\nu+2)(n_p+\nu)}\right)-C_1(\nu)\right\}=\frac{1}{4}D, \qquad \text{(A3.16)}$$

where $C(\nu)=\Psi(\frac{n_p+\nu}{2})-\Psi(\frac{\nu}{2})$ and $\Psi(.)$ is the digamma function.

Finally, using the results (A3.3), (A3.6),(A3.8) (A3.12),(A3.14) and (A3.16), we have therefore that

$$I_1(\theta^*)=\begin{bmatrix} \frac{1}{2\sigma^4}B & \frac{1}{4\sigma^2}B_{11} & \frac{1}{\sigma^2}B_{12} \\ \frac{1}{4\sigma^2}B_{11} & \frac{1}{4}C & C_{11} \\ \frac{1}{\sigma^2}B_{12} & C_{11} & \frac{1}{4}D \end{bmatrix} \qquad \text{(A3.17)}$$

and

$$\pi(\boldsymbol{\beta},\sigma^2,\phi,\nu)\propto\frac{1}{\sigma^2}\left(BCD+16B_{11}C_{11}B_{12}-8B_{12}^2C-16BC_{11}^2-\frac{1}{2}B_{11}^2D\right)^{\frac{1}{2}}.$$

A4 Proof of $|I_1(\theta^*)|>0$

Through elemental row operations $R_2:R_3-\frac{\sigma^2}{2}\frac{B_{11}}{B}R_1$ and $R_3:R_3-\frac{2\sigma^2B_{12}}{B}R_1$, we obtain a reduced form $R_1(\theta^*)$ for the matrix $I_1(\theta^*)$ in (A3.17), which is given by

$$\begin{aligned} R_1(\theta^*) &= \begin{bmatrix} \frac{1}{2\sigma^4}B & \frac{1}{4\sigma^2}B_{11} & \frac{1}{\sigma^2}B_{12} \\ 0 & \frac{1}{4}C-\frac{1}{8}\frac{B_{11}^2}{B} & C_{11}-\frac{1}{2}\frac{B_{12}B_{11}}{B} \\ 0 & C_{11}-\frac{1}{2}\frac{B_{12}B_{11}}{B} & \frac{1}{4}D-\frac{2B_{12}^2}{B} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2\sigma^4}B & \frac{1}{4\sigma^2}B_{11} & \frac{1}{\sigma^2}B_{12} \\ 0 & \frac{1}{4}C-\frac{1}{8}\frac{B_{11}^2}{B} & 0 \\ 0 & 0 & \frac{1}{4}D-\frac{2B_{12}^2}{B} \end{bmatrix}, \end{aligned}$$

once $C_{11}-\frac{1}{2}\frac{B_{12}B_{11}}{B}=0$. Then, we obtain an upper triangular matrix and we have that $|I_1(\theta^*)|>0$ if $|R_1(\theta^*)|>0$. So, we only need to show that the product of the diagonal elements of the matrix $R_1(\theta^*)$ is greater than 0. First, note that $\frac{1}{2\sigma^4}B>0$ and it can be shown that

$$C>\left(\frac{2(\nu+4)}{(\nu-4)(\nu-2)}+\frac{n_p\nu}{(\nu+n_p+2)(\nu-2)(\nu-4)}\right)tr^2\left[\frac{\partial\mathbf{R}}{\partial\phi}\mathbf{R}^{-1}\mathbf{P}\right],$$

then

$$\begin{aligned}\frac{1}{4}C - \frac{1}{8}\frac{B_{11}^2}{B} &> \frac{1}{2(\nu-2)}\left(\frac{\nu+4}{(\nu-4)} + \frac{n_p\nu(6-\nu)}{(\nu-4)(\nu-2)(\nu+\nu+2))}\right) tr^2\left[\frac{\partial \mathbf{R}}{\partial \phi}\mathbf{R}^{-1}\mathbf{P}\right] \\ &> 0,\end{aligned}$$

since $\left(\frac{n_p\nu(6-\nu)}{(\nu-4)(\nu-2)(\nu+n_p+2))}\right) \geqslant 0$ when $4 < \nu \leqslant 6$, and less than 1 (in absolute value) when $\nu > 6$. Now, for $n_p > 0$ and $\nu > 4$,

$$\begin{aligned}\frac{1}{4}D - \frac{2B_{12}^2}{B} &= \frac{1}{4}D - 2\left(\frac{\nu+n_p+2}{n_p\nu}\right)\frac{n_p^2}{(n_p+\nu)^2(n_p+\nu+2)^2} \\ &= \frac{1}{4}D - \frac{2n_p}{(n_p+\nu)^2(n_p+\nu+2)\nu} \\ &= -\frac{n_p}{(n_p+\nu)(n_p+\nu+2)\nu}\left(\frac{n_p+\nu+4}{2} + \frac{2}{(n_p+\nu)}\right) - \frac{1}{4}C_1(\nu) > 0.\end{aligned}$$

This result was also verified trought a bivariate plot of $n_p > 0$ and $\nu > 4$. Then, $|I_1(\theta^*)| > 0$.

APPENDIX B – Appendix for chapter 3

B1 Complementary results in the simulation study: Student-t process

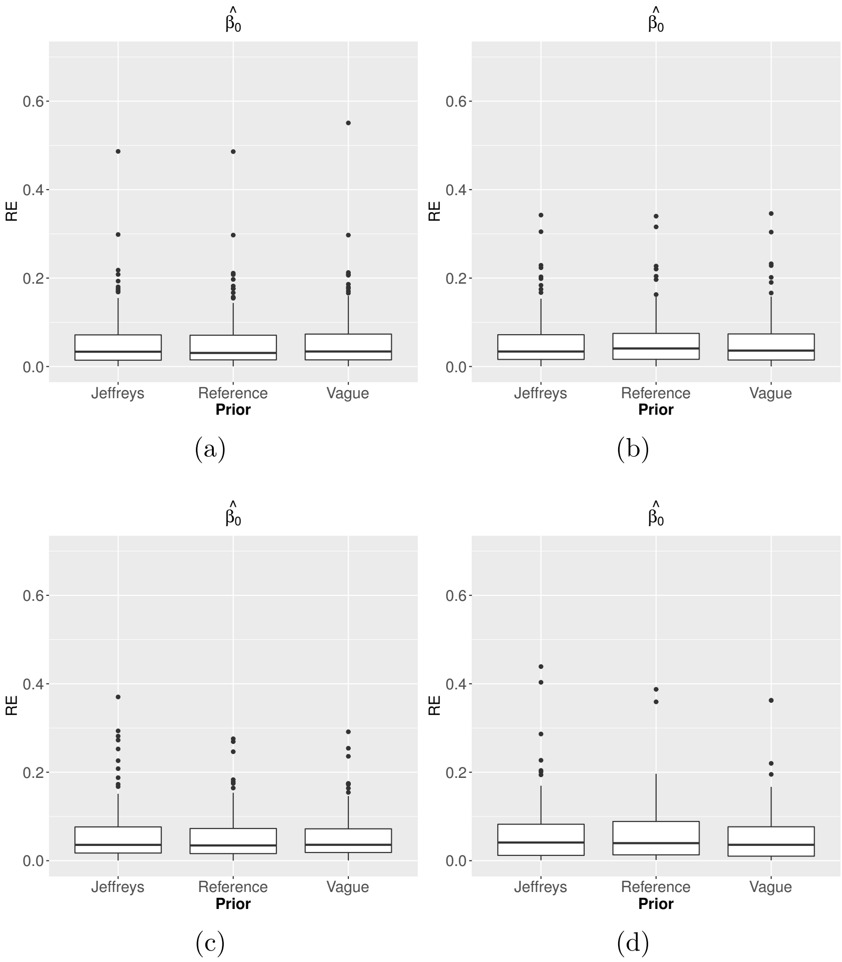


Figure 7 – Simulation study 1. Student-t process under **Scenario 1**, relative error for $\hat{\beta}_0$ considering (a) $n = 50$, (b) $n = 100$ (c) $n = 250$ (d) $n = 500$.

Table 15 – Simulation study 1. Coverage probability (and expected log-length) for the mean structure parameter β_0 (**Scenario 1**) considering different sample sizes.

Prior/n	50	100	250	500
reference	0.95 (0.42)	0.96 (0.39)	0.97 (0.35)	0.97 (0.30)
vague	0.91 (0.39)	0.95 (0.35)	0.94 (0.36)	0.96 (0.33)
Jef rul	0.73 (0.24)	0.82 (0.22)	0.828 (0.18)	0.84 (0.19)

Table 16 – Simulation study 1. Coverage probability (and expected log-length) for the mean structure parameters under **Scenario 2** considering different sample sizes.

n	Prior	β_0	β_1	β_2	β_3	β_4	β_5
50	*reference*	0.94 (1.52)	0.929 (0.20)	0.91 (0.21)	0.88 (-2.24)	0.94 (-2.25)	0.93 (-2.30)
	vague	0.92 (1.49)	0.92 (0.18)	0.93 (0.2)	0.9 (-2.28)	0.93 (-2.28)	0.92(-2.31)
	Jef rul	0.79 (1.21)	0.78 (-0.05)	0.86 (-0.03)	0.78 (-2.50)	0.87 (-2.48)	0.79 (-2.56)
	Jef ind		0.89 (-0.13)	0.93 (-0.14)	0.88 (-2.37)	0.93 (-2.37)	0.95 (-2.57)
100	*reference*	0.98 (1.45)	0.97 (0.14)	0.93 (0.14)	0.93 (-2.31)	0.92 (-2.31)	0.97 (-2.36)
	vague	0.94 (1.39)	0.97 (0.10)	0.89 (0.10)	0.91 (-2.36)	0.89 (-2.36)	0.97 (-2.41)
	Jef rul	0.86 (1.17)	0.87 (-0.10)	0.84(-0.10)	0.88 (-2.55)	0.84 (-2.55)	0.92 (-2.60)
	Jef ind		0.89 (-0.14)	0.88 (-0.14)	0.89 (-2.40)	0.87 (-2.41)	0.93 (-2.59)
250	*reference*	0.95 (1.39)	0.97 (0.08)	0.93 (0.07)	0.95 (-2.39)	0.95 (-2.39)	0.95 (-2.43)
	vague	0.91 (1.32)	0.96 (0.03)	0.91 (0.02)	0.93 (-2.43)	0.89 (-2.43)	0.94 (-2.48)
	Jef rul	0.78 (1.09)	0.80 (-0.12)	0.77 (-0.12)	0.79 (-2.60)	0.82 (-2.62)	0.77 (-2.61)
	Jef ind		0.81 (-0.13)	0.85 (-0.13)	0.83 (-2.42)	0.82 (-2.41)	0.82 (-2.52)
500	*reference*	0.88 (1.31)	0.89 (-0.15)	0.86 (-0.14)	0.875 (-2.38)	0.875 (-2.35)	0.92 (-2.50)
	vague	0.91 (1.30)	0.96 (-0.02)	0.91 (-0.00)	0.93 (-2.48)	0.89 (-2.47)	0.94 (-2.50)
	Jef rul	0.76 (1)	0.77 (-0.10)	0.79 (-0.10)	0.78 (-2.55)	0.83 (-2.55)	0.84 (-2.60)
	Jef ind		0.78 (-0.14)	0.76 (-0.14)	0.72 (-2.40)	0.78 (-2.41)	0.85 (-2.59)

Table 17 – Simulation study 1. Coverage probability (and expected log-length) for the variance structure parameters under two simulation Scenarios and considering different sample sizes.

n	Scenario	Prior	σ^2	ϕ	ν
$n = 50$	Scenario 1	*reference*	0.92 (0.95)	0.96 (0.99)	1 (2.56)
		vague	0.37 (-0.05)	0.80 (0.88)	0.79 (2.42)
		Jef rul	0.45 (1.11)	0.62 (0.85)	0.68 (2.63)
	Scenario 2	*reference*	0.94 (0.83)	0.97 (1.12)	1 (2.56)
		vague	0.31 (-0.1)	0.90 (0.97)	0 (2.42)
		Jef rul	0.22 (-0.55)	0.56 (0.6)	0.06 (2.51)
		Jef ind	0.43 (0.28)	0.85 (0.87)	0.08 (2.55)
$n = 250$	Scenario 1	*reference*	0.95 (0.89)	0.96 (0.98)	1 (2.39)
		vague	0.36 (-0.12)	0.88 (0.76)	0.85 (2.38)
		Jef rul	0.45 (0.54)	0.87 (0.66)	0.79 (2.35)
	Scenario 2	*reference*	0.96 (0.99)	0.96 (1.07)	1 (2.56)
		vague	0.31 (-0.24)	0.92 (0.82)	0.22 (2.35)
		Jef rul	0.55 (-0.33)	0.6 (0.58)	0 (2.32)
		Jef ind	0.50 (0.52)	0.87 (1.03)	0.14 (2.52)
$n = 500$	Scenario 1	*reference*	0.95 (0.82)	0.96 (0.95)	1 (2.36)
		vague	0.37 (-0.14)	0.91 (0.69)	0.94 (2.36)
		Jef rul	0.50 (0.4)	0.88 (0.87)	0.9 (2.36)
	Scenario 2	*reference*	0.95 (0.91)	0.90 (1.05)	1 (2.36)
		vague	0.37 (-0.13)	0.827 (0.85)	0.225 (2.38)
		Jef rul	0.56 (-0.88)	0.66 (0.61)	0 (2.32)
		Jef ind	0.5 (0.55)	0.704 (0.99)	0.16 (2.5)

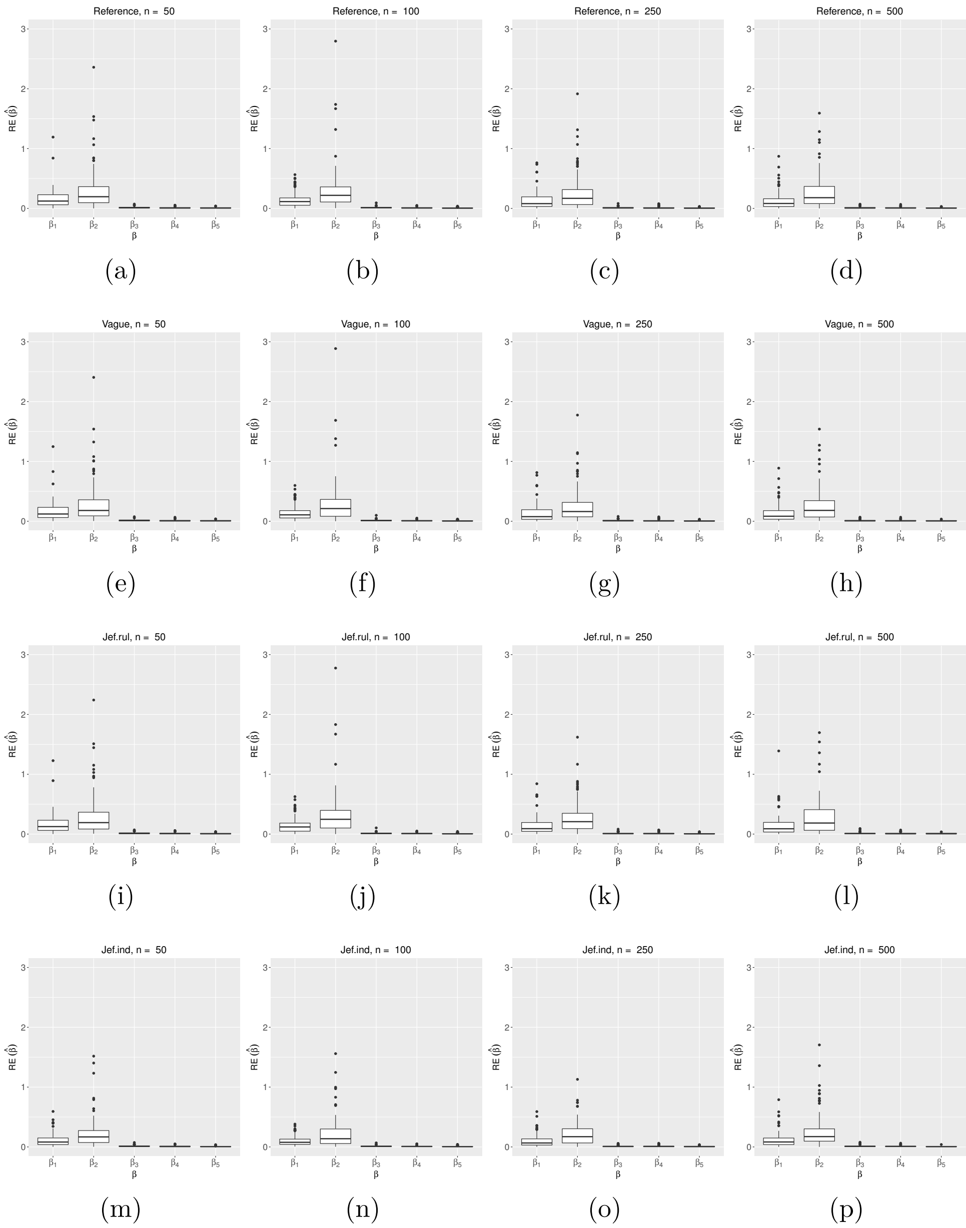


Figure 8 – Simulation study 1. Student-t process under **Scenario 2**, relative error for $\hat{\boldsymbol{\beta}}$ considering different sample sizes.

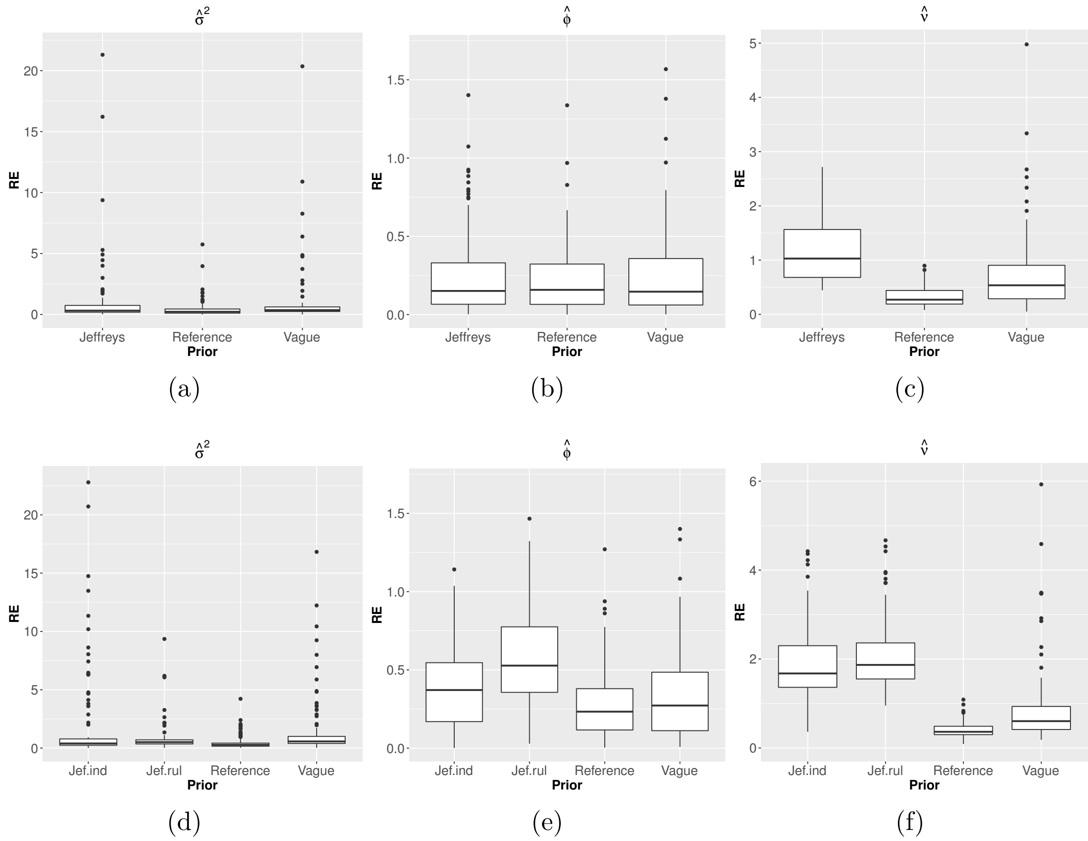


Figure 9 – Simulation study 1. Relative error for $\hat{\sigma}$, $\hat{\phi}$ and $\hat{\nu}$ under **Scenario 1** (figures (a)-(c)) and **Scenario 2** (figures (d)-(f)) for $n = 50$.

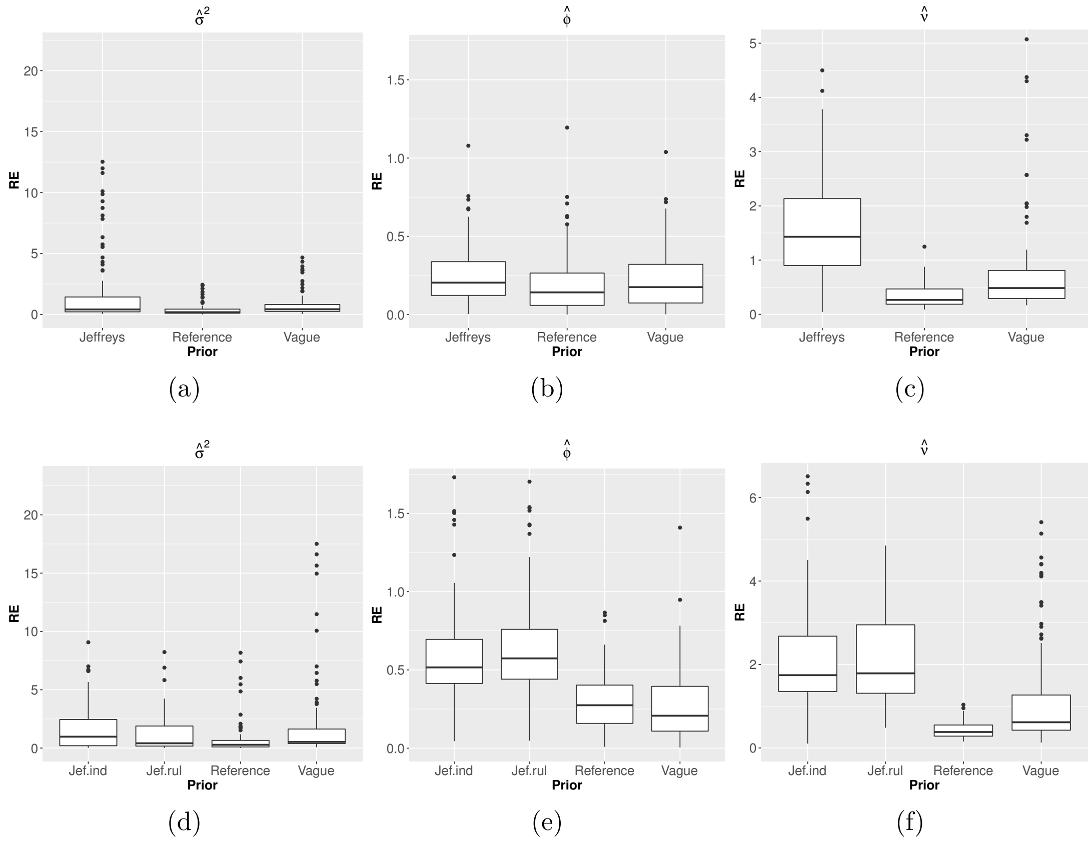


Figure 10 – Simulation study 1. Relative error for $\hat{\sigma}$, $\hat{\phi}$ and $\hat{\nu}$ under **Scenario 1** (figures (a)-(c)) and **Scenario 2** (figures (d)-(f)) for $n = 250$.

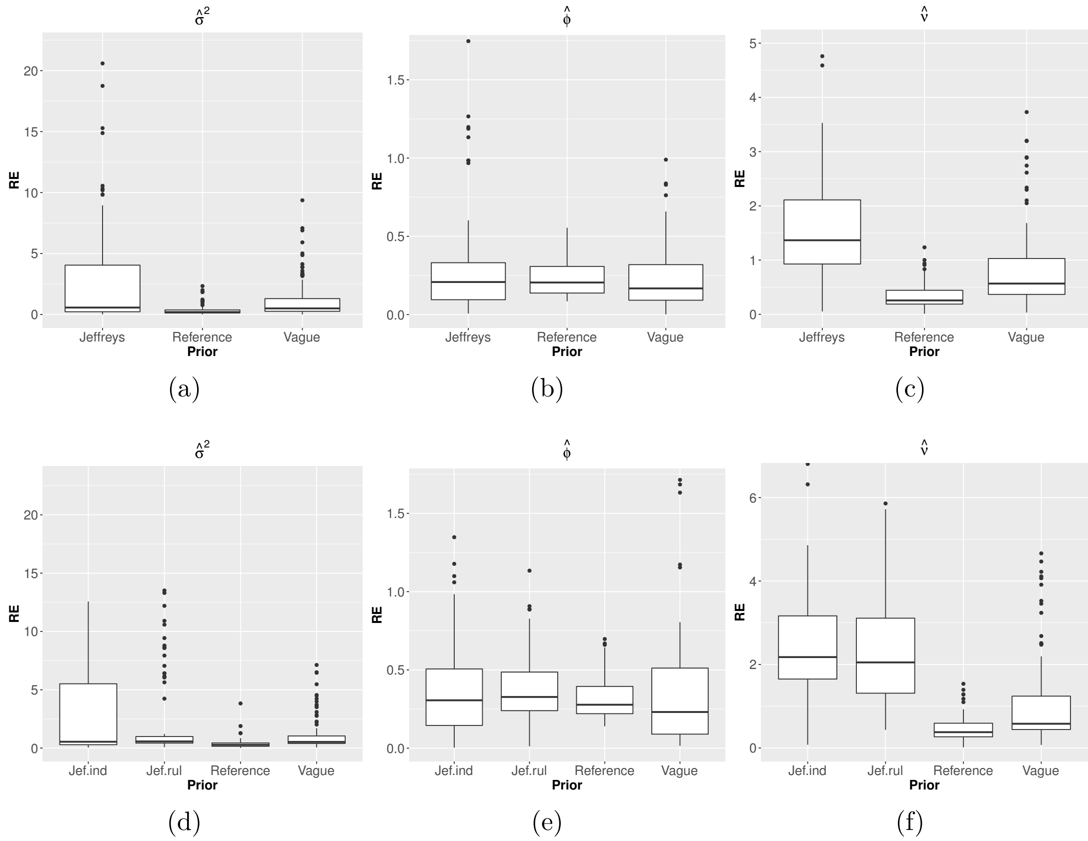


Figure 11 – Simulation study 1. Relative error for $\hat{\sigma}$, $\hat{\phi}$ and $\hat{\nu}$ under **Scenario 1** (figures (a)-(c)) and **Scenario 2** (figures (d)-(f)) for $n = 500$.

B2 Complementary results: Student-t versus Gaussian spatial regression

Table 18 – Simulation study 2. Expected Relative Error (and standard deviation) and Coverage Probability (and expected log-length) for the mean structure parameter β_0 (**Scenario 1**), considering different distributions and sample sizes.

Distribution	MRE-CP	Prior / n	50	100	250	500
T-SR	MRE	*reference*	0.07 (0.07)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
		vague	0.07 (0.07)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
		Jef rul	0.07 (0.07)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
	CP	*reference*	0.94 (0.56)	0.97 (0.48)	0.94 (0.51)	0.84 (0.06)
		vague	0.95 (0.73)	0.99 (0.65)	0.97 (0.67)	0.82 (0.04)
		Jef rul	0.86 (0.41)	0.92 (0.36)	0.93 (0.36)	0.75 (-0.08)
N-SR	MRE	*reference*	0.07 (0.07)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
		vague	0.07 (0.07)	0.06 (0.06)	0.05 (0.05)	0.05 (0.05)
		Jef rul	0.07 (0.07)	0.06 (0.06)	0.05(0.05)	0.05 (0.05)
	CP	*reference*	0.95 (0.52)	0.97 (0.48)	0.98 (0.49)	0.85 (0.06)
		vague	0.92 (0.43)	0.96 (0.41)	0.96 (0.43)	0.94 (0.42)
		Jef rul	0.86 (0.35)	0.96 (0.35)	0.70 (-0.11)	0.69 (-0.07)

Table 19 – Simulation study 2. Expected relative error (and standard deviation) for the mean structure parameters under **Scenario 2**, considering $n = 50, 100$ and different distributions.

n	Process	Prior	β_1	β_2	β_3	β_4	β_5
50	T-SR	*reference*	0.17 (0.17)	0.37 (0.32)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		vague	0.17 (0.17)	0.37 (0.34)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.17 (0.17)	0.38 (0.3)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.11 (0.13)	0.32 (0.29)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
	N-SR	*reference*	0.17 (0.17)	0.37 (0.32)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		vague	0.17 (0.17)	0.37 (0.32)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.17 (0.17)	0.38 (.32)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.11 (0.13)	0.30 (0.28)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
100	T-SR	*reference*	0.11 (0.16)	0.26 (0.29)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
		vague	0.11 (0.16)	0.27 (0.29)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.12 (0.16)	0.28 (0.29)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.09 (0.12)	0.21 (0.26)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
	N-SR	*reference*	0.11 (0.16)	0.28 (0.3)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
		vague	0.11 (0.16)	0.27 (0.29)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.12 (0.16)	0.28 (0.3)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.09 (0.12)	0.21 (0.25)	0.01 (0.02)	0.01 (0.01)	0.01 (0.01)

Table 20 – Simulation study 2. Expected relative error (and standard deviation) for the mean structure parameters under **Scenario 2**, considering $n = 250, 500$ and different distributions.

n	Distribution	Prior	β_1	β_2	β_3	β_4	β_5
250	T-SR	*reference*	0.11 (0.11)	0.26 (0.31)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		vague	0.11 (0.11)	0.26 (0.29)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.11 (0.12)	0.28 (0.33)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.08 (0.08)	0.19 (0.17)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
	N-SR	*reference*	0.11 (0.11)	0.26 (0.29)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		vague	0.11 (0.12)	0.28 (0.34)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.13 (0.15)	0.32 (0.43)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.11 (0.13)	0.26 (0.31)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
500	T-SR	*reference*	0.12 (0.13)	0.23 (0.22)	0.01 (0.01)	0.01 (0.01)	0.01(0.01)
		vague	0.12 (0.13)	0.24 (0.22)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.15 (0.18)	0.32 (0.34)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.13 (0.16)	0.28 (0.36)	0.01 (0.01)	0.01 (0.02)	0.01 (0.01)
	N-SR	*reference*	0.12 (0.13)	0.24 (0.22)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		vague	0.12 (0.13)	0.24 (0.21)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef rul	0.12 (0.13)	0.25 (0.22)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
		Jef ind	0.10 (0.12)	0.23 (0.21)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)

Table 21 – Simulation study 2. Coverage probability (and expected log-length) for the mean structure parameters under **Scenario 2**, considering $n = 50, 100$ and different distributions.

n	Distribution	Prior	β_0	β_1	β_2	β_3	β_4	β_5
n = 50	T-SR	*reference*	0.90 (1.56)	0.9 (0.29)	0.95 (0.28)	0.94 (-2.15)	0.92 (-2.15)	0.89 (-2.19)
		vague	0.88 (1.49)	0.88 (0.24)	0.87 (0.23)	0.88 (-2.20)	0.89 (-2.20)	0.85 (-2.24)
		Jef rul	0.73 (1.26)	0.82 (0.07)	0.77 (0.08)	0.85 (-2.39)	0.81 (-2.39)	0.75 (-2.45)
		Jef ind		0.92 (0.10)	0.91 (0.12)	0.96 (-2.22)	0.94 (-2.23)	0.96 (-2.40)
	N-SR	*reference*	0.859 (1.47)	0.87 (0.19)	0.84 (0.208)	0.89 (-2.23)	0.89 (-2.24)	0.85 (-2.29)
		vague	0.91 (1.67)	0.91 (0.35)	0.89 (0.35)	0.90 (-2.10)	0.93 (-2.10)	0.88 (-2.12)
		Jef rul	0.69 (1.15)	0.78 (-0.06)	0.72 (-0.06)	0.82 (-2.48)	0.78 (-2.49)	0.75 (-2.56)
		Jef ind		0.96 (0.06)	0.90 (0.07)	0.99 (-2.18)	0.91 (-2.18)	0.94 (-2.34)
n = 100	T-SR	*reference*	0.93 (1.33)	0.96 (0.12)	0.94 (0.13)	0.92 (-2.33)	0.93 (-2.32)	0.94 (-2.38)
		vague	0.94 (1.37)	0.95 (0.08)	0.91 (0.08)	0.92 (-2.37)	0.92 (-2.37)	0.95 (-2.41)
		Jef rul	0.88 (1.15)	0.91 (-0.10)	0.81 (-0.11)	0.87 (-2.54)	0.87 (-2.54)	0.85 (-2.60)
		Jef ind		0.94 (-0.15)	0.92 (-0.16)	0.91 (-2.40)	0.95 (-2.41)	0.91 (-2.58)
	N-SR	*reference*	0.94 (1.34)	0.95 (0.05)	0.90 (0.05)	0.91 (-2.41)	0.91 (-2.40)	0.92 (-2.44)
		vague	1 (1.56)	0.96 (0.19)	0.91 (0.20)	0.94 (-2.27)	0.94 (-2.27)	0.97 (-2.30)
		Jef rul	0.85 (1.03)	0.86 (-0.22)	0.79 (-0.22)	0.83 (-2.65)	0.84 (-2.27)	0.80 (-2.72)
		Jef ind		0.95 (-0.1)	0.93 (-0.10)	0.94 (-2.38)	0.96 (-2.38)	0.93 (-2.54)

Table 22 – Simulation study 2. Coverage probability (and expected log-length) for the mean structure parameters under **Scenario 2**, considering $n = 250, 500$ and different distributions.

n	Distribution	Prior	β_0	β_1	β_2	β_3	β_4	β_5
n = 250	T-SR	*reference*	0.97 (1.44)	0.96 (0.11)	0.93 (0.127)	0.94 (-2.34)	0.95 (-2.34)	0.96 (-2.40)
		vague	0.96 (1.40)	0.96 (0.08)	0.92 (0.10)	0.98 (-2.37)	0.93 (-2.36)	0.93 (-2.43)
		Jef rul	0.81 (0.40)	0.83 (-0.96)	0.84 (-0.86)	0.78 (-3.34)	0.81 (-3.27)	0.89 (-3.41)
		Jef ind		0.62 (-0.87)	0.65 (-0.81)	0.61 (-3.14)	0.62 (-3.06)	0.71 (-3.23)
	N-SR	*reference*	0.95 (1.40)	0.93 (0.07)	0.94 (0.07)	0.93 (-2.40)	0.923 (-2.40)	0.96 (-2.43)
		vague	0.97 (1.59)	0.97 (0.21)	0.95 (0.22)	0.96 (-2.26)	0.94 (-2.27)	0.98 (-2.30)
		Jef rul	0.86 (1.07)	0.89 (-0.20)	0.82 (-0.18)	0.86 (-2.64)	0.85 (-2.64)	0.91 (-2.71)
		Jef ind		0.95 (-0.06)	0.94 (-0.083)	0.94 (-2.37)	0.967 (-2.38)	0.96 (-2.55)
n = 500	T-SR	*reference*	0.9 (1)	0.88 (-0.25)	0.84 (-0.24)	0.87 (-2.7)	0.87 (-2.69)	0.89 (-2.77)
		vague	0.96 (1.318)	0.94 (0.02)	0.94 (0.01)	0.94 (-2.44)	0.89 (-2.45)	0.82 (-2.50)
		Jef rul	0.73 (0.32)	0.78 (-0.99)	0.832 (-0.94)	0.86 (-3.40)	0.84 (-3.41)	0.88 (-3.47)
		Jef ind		0.68 (-0.82)	0.614 (-0.78)	0.65 (-3.1)	0.62 (-3.0)	0.68 (-3.31)
	N-SR	*reference*	0.88 (0.99)	0.88 (-0.26)	0.82 (-0.26)	0.85 (-2.71)	0.85 (-2.72)	0.88 (-2.79)
		vague	0.86 (1)	0.88 (-0.25)	0.80 (-0.25)	0.88 (-2.70)	0.86 (-2.7)	0.89 (-2.77)
		Jef rul	0.83 (0.89)	0.86 (-0.35)	0.75 (-0.35)	0.84 (-2.79)	0.78 (-2.8)	0.841 (-2.87)
		Jef ind		0.85 (-0.52)	0.77 (-0.52)	0.83 (-2.75)	0.79 (-2.75)	0.88 (-2.98)

Table 23 – Simulation study 2. Expected relative error (and standard deviation) of the variance parameters for both Student-t and normal distributions considering $n = 50$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.41 (0.55)	0.19 (0.19)	0.31 (0.16)
		vague	0.88 (1.45)	0.20 (0.19)	0.76 (0.73)
		Jef rul	1.02 (2.05)	0.23 (0.21)	1.15 (0.60)
	N-SR	*reference*	0.82 (1.50)	0.274 (0.261)	-
		vague	1.48 (3.14)	0.353 (0.347)	-
		Jef rul	2.41 (7.34)	0.265 (0.233)	-
Scenario 2	T-SR	*reference*	0.84 (0.75)	0.26 (0.2)	0.369 (0.3)
		vague	1.91 (4.07)	0.25 (0.31)	3.50 (7.5)
		Jef rul	1.74 (4.9)	0.57 (0.34)	2.09 (0.83)
		Jef ind	3.31 (7.72)	0.34 (0.34)	1.845 (0.87)
	N-SR	*reference*	1.24 (1.77)	0.36 (0.22)	-
		vague	2.39 (5.82)	0.43 (0.18)	-
		Jef rul	1.98 (5.33)	0.60 (0.27)	-
		Jef ind	4.12 (9.37)	0.37 (0.24)	-

Table 24 – Simulation study 2. Expected relative error (and standard deviation) of the variance parameters for both Student-t and normal distributions considering $n = 250$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.425 (0.53)	0.20 (0.19)	0.33 (0.19)
		vague	0.96 (1.34)	0.23 (0.29)	1.09 (1.3)
		Jef rul	2.62 (3.94)	0.28 (0.34)	1.7 (0.97)
	N-SR	*reference*	1.06 (2.85)	0.289 (0.257)	-
		vague	1.75 (6.10)	0.367 (0.345)	-
		Jef rul	3.45 (8.78)	0.260 (0.228)	-
Scenario 2	T-SR	*reference*	0.69 (1.24)	0.31 (0.25)	0.42 (0.35)
		vague	1.86 (3.91)	0.27 (0.40)	5.64 (7.02)
		Jef rul	3.13 (6.76)	0.40 (0.30)	2.26 (1.34)
		Jef ind	4.15 (7.24)	0.42 (0.35)	2.34 (1.45)
	N-SR	*reference*	1.77 (3.74)	0.27 (0.19)	-
		vague	3.29 (8.52)	0.44 (0.16)	-
		Jef rul	3.24 (7.49)	0.4 (0.33)	-
		Jef ind	4.89 (7.95)	0.3 (0.25)	-

Table 25 – Simulation study 2. Expected relative error (and standard deviation) of the variance parameters for both Student-t and normal distributions considering $n = 500$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.41 (0.55)	0.19 (0.19)	0.31 (0.16)
		vague	0.88 (1.45)	0.20 (0.19)	0.76 (0.73)
		Jef rul	1.02 (2.05)	0.23 (0.21)	1.15 (0.60)
	NSR	*reference*	0.36 (0.59)	0.24 (0.15)	-
		vague	1.24 (2.03)	0.33 (0.144)	-
		Jef rul	2.84 (5.17)	0.26 (0.156)	-
Scenario 2	T-SR	*reference*	0.84 (0.75)	0.26 (0.2)	0.369 (0.3)
		vague	1.91 (4.07)	0.25 (0.31)	3.50 (7.5)
		Jef rul	1.74 (4.9)	0.57 (0.34)	2.09 (0.83)
		Jef ind	3.31 (7.72)	0.34 (0.34)	1.84 (0.87)
	N-SR	*reference*	0.49 (0.70)	0.35 (0.3)	-
		vague	1.38 (2.23)	0.50 (0.35)	-
		Jef rul	2.45 (4.48)	0.67 (0.32)	-
		Jef ind	3.10 (5.50)	0.4 (0.36)	-

Table 26 – Simulation study 2. Coverage probability (and expected log-length) of the variance parameters for both Student-t and normal distributions considering $n = 50$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.97 (0.89)	0.93 (1.03)	1 (2.55)
		vague	0.42 (-0.01)	0.93 (0.90)	0.92 (2.45)
		Jef rul	0.56 (0.35)	0.94 (0.90)	0.97 (2.63)
	N-SR	*reference*	1.00 (1.35)	0.990 (1,376)	-
		vague	0.47 (0.18)	0.949 (1,544)	-
		Jef rul	0.54 (0.61)	0.929 (1,173)	-
Scenario 2	T-SR	*reference*	0.95 (0.99)	0.93 (1.07)	1 (2.55)
		vague	0.33 (0.01)	0.86 (0.97)	0.06 (2.41)
		Jef rul	0.27 (-0.32)	0.58 (0.63)	0 (2.50)
		Jef ind	0.46 (0.52)	0.88 (0.52)	0.08 (0.52)
	N-SR	*reference*	0.97 (1.39)	0.91 (1.25)	-
		vague	0.51 (0.20)	1.00 (1.71)	-
		jef rul	0.34 (-0.30)	0.31 (0.32)	-
		jef ind	0.56 (0.70)	0.97 (1.53)	-

Table 27 – Simulation study 2. Coverage probability (and expected log-length) of the variance parameters for both Student-t and normal distributions considering $n = 250$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.97 (0.53)	0.93 (0.96)	0.95 (2.55)
		vague	0.41 (1.34)	0.92 (0.72)	0.93 (2.40)
		Jef rul	0.45 (3.94)	0.71 (0.64)	0.74 (2.51)
	N-SR	*reference*	0.98 (1.53)	0.918 (1,123)	-
		vague	0.60 (0.31)	0.938 (1,37)	-
		Jef rul	0.61 (1.05)	0.866 (0,946)	-
Scenario 2	T-SR	*reference*	0.97 (1.08)	0.97 (1.07)	1 (2.56)
		vague	0.41 (0.14)	0.81 (0.79)	0.19 (2.34)
		Jef rul	0.28 (-0.44)	0.51 (0.24)	0 (2.27)
		Jef ind	0.47 (0.64)	0.68 (0.65)	0.12 (0.64)
	N-SR	*reference*	0.97 (1.53)	0.95 (1.19)	-
		vague	0.49 (0.30)	0.94 (1.59)	-
		Jef rul	0.16 (-0.65)	0.45 (0.24)	-
		Jef ind	0.57 (0.75)	0.96 (1.40)	-

Table 28 – Simulation study 2. Coverage probability (and expected log-length) of the variance parameters for both Student-t and normal distributions considering $n = 500$.

Scenario	Distribution	Prior	σ^2	ϕ	ν
Scenario 1	T-SR	*reference*	0.85 (0.47)	0.85 (0.89)	0.98 (2.47)
		vague	0.4 (0.09)	0.83 (0.77)	0.83 (2.43)
		Jef rul	0.45 (0.67)	0.69 (0.68)	0.71 (2.57)
	N-SR	*reference*	1 (0.89)	0.75 (0.66)	-
		vague	0.54 (0.34)	0.73 (0.71)	-
		Jef rul	0.55 (0.75)	0.79 (0.62)	-
Scenario 2	T-SR	*reference*	0.82 (1.05)	0.88 (0.89)	1 (2.54)
		vague	0.35 (0.11)	0.83 (0.72)	0.22 (2.33)
		Jef rul	0.17 (-0.41)	0.52 (0.65)	0 (2.31)
		Jef ind	0.50 (0.62)	0.62 (0.64)	0.14 (0.45)
	N-SR	*reference*	0.89 (0.78)	0.85 (0.93)	-
		vague	0.46 (0.18)	0.85 (0.88)	-
		Jef rul	0.50 (-0.77)	0.81 (0.74)	-
		Jef ind	0.5 (0.47)	0.86 (0.78)	-

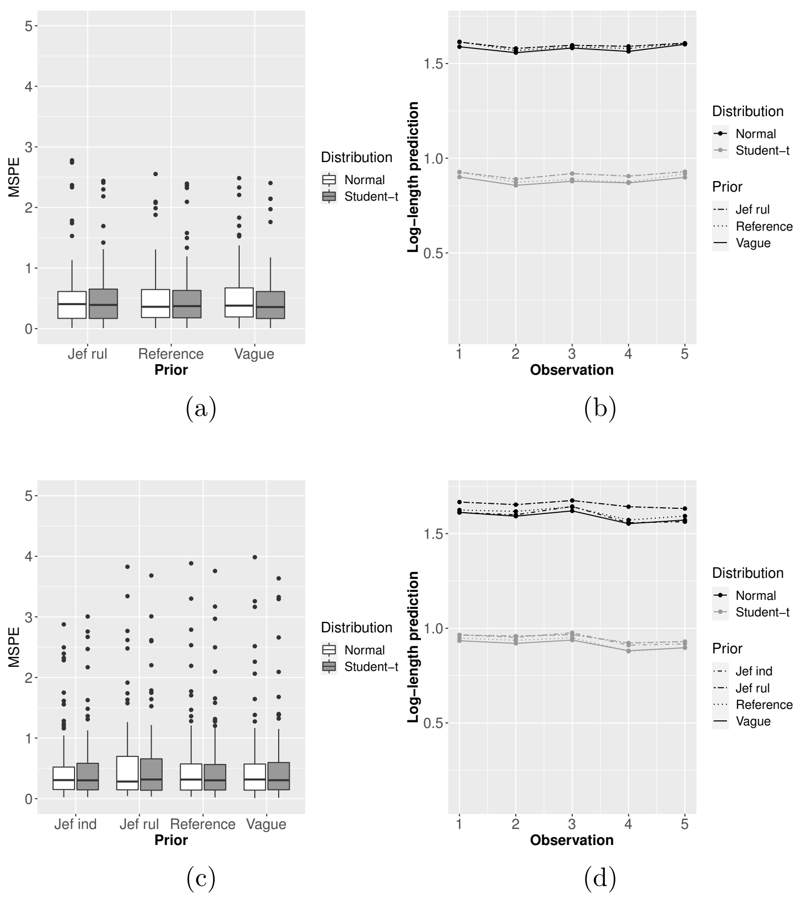


Figure 12 – Simulation study 2. MSPE and log-length of prediction points under **Scenario 1** (figures (a)-(b)) and **Scenario 2** (figures (c)-(d)) for $n = 50$.

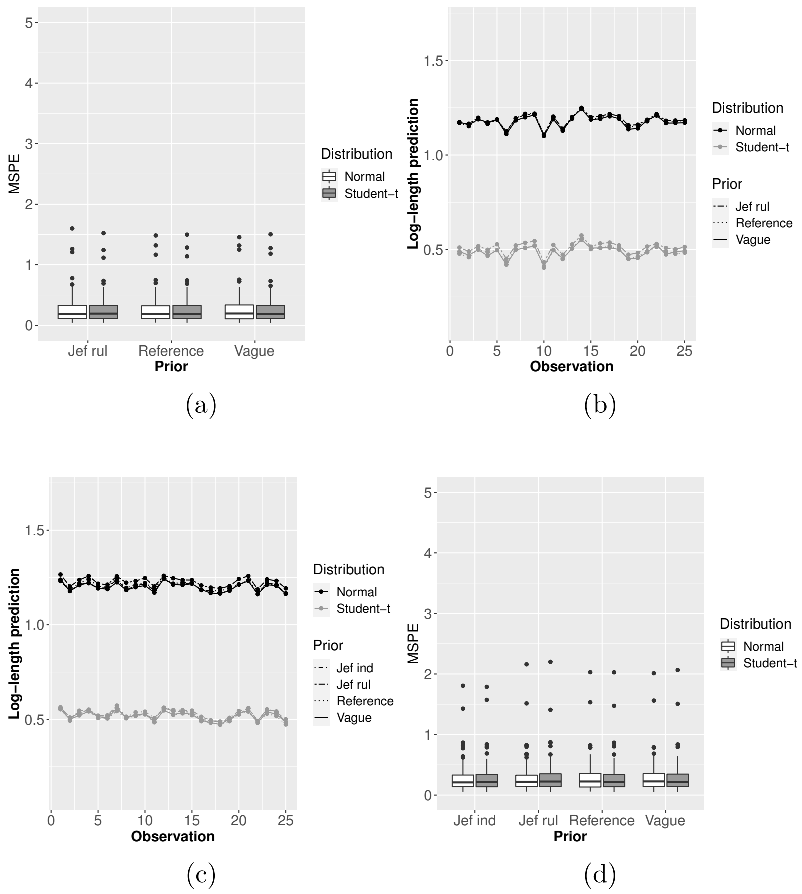
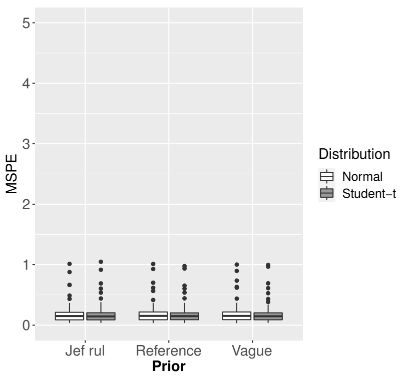
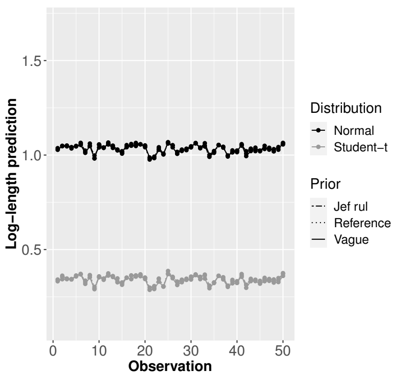


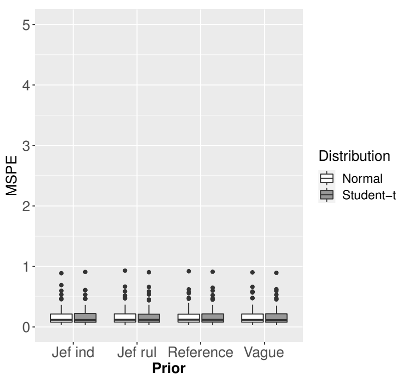
Figure 13 – Simulation study 2. MSPE and log-length of prediction points under **Scenario 1** (figures (a)-(b)) and **Scenario 2** (figures (c)-(d)) for $n = 250$.



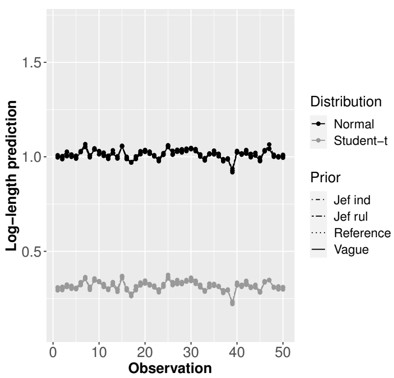
(a)



(b)



(c)



(d)

Figure 14 – Simulation study 2. MSPE and log-length of prediction points under **Scenario 1** (figures (a)-(b)) and **Scenario 2** (figures (c)-(d)) for $n = 500$.

APPENDIX C – Appendix for chapter 5

C1 Further simulations results

In this section results about the $\boldsymbol{\beta}$ parameters is presented under different links, sample size and prior specification.

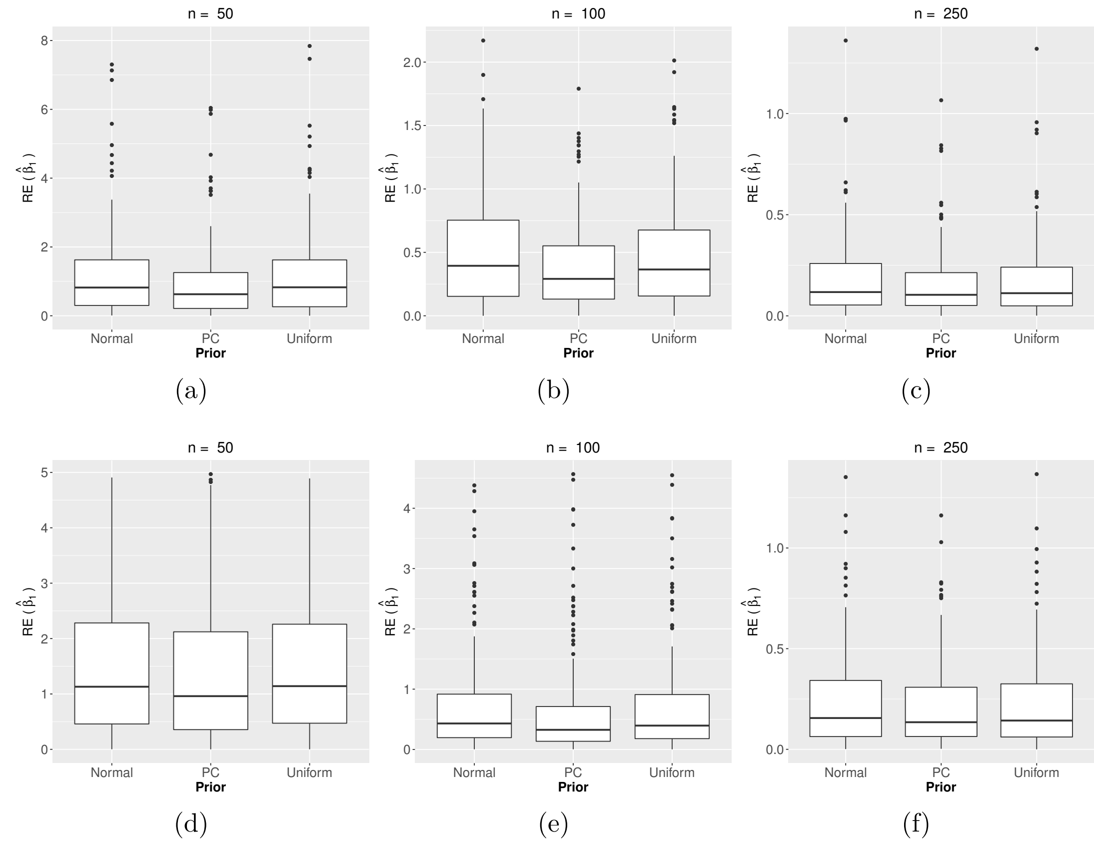


Figure 15 – Simulation study. Relative error for $\hat{\beta}_1$ under the GL link ((a)-(c)) and the PP link ((d)-(f)) and different sample sizes.

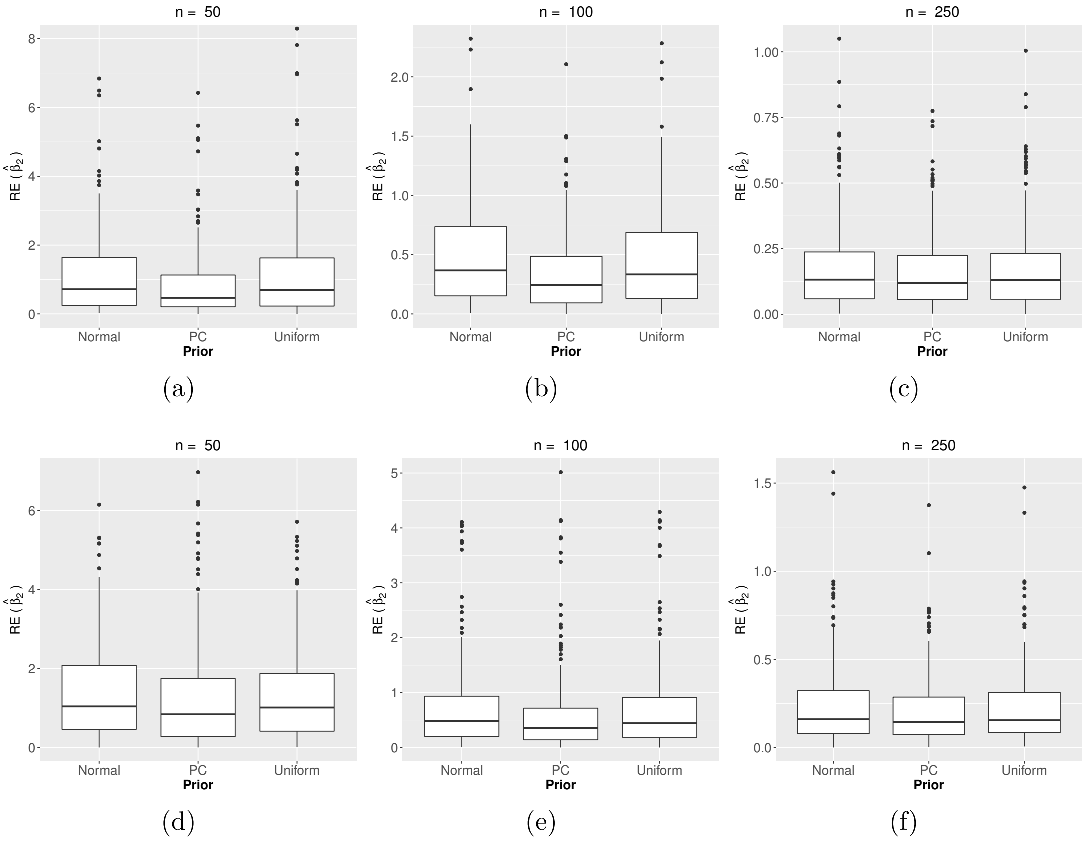


Figure 16 – Simulation Study. Relative error for $\hat{\beta}_2$ under the GL link ((a)-(c)) and the PN link ((d)-(f)) and different sample sizes.

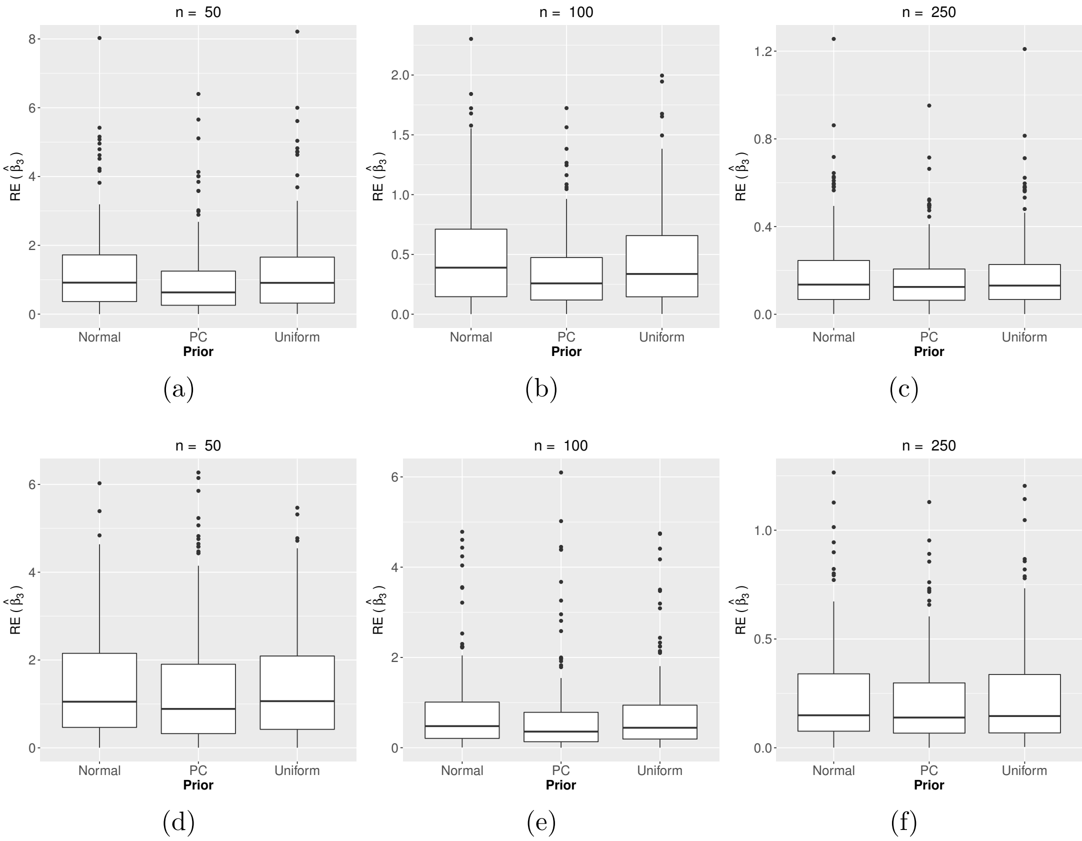
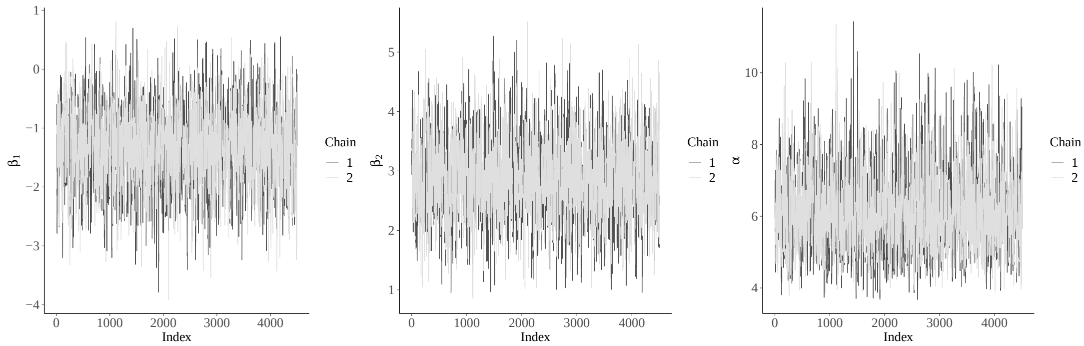


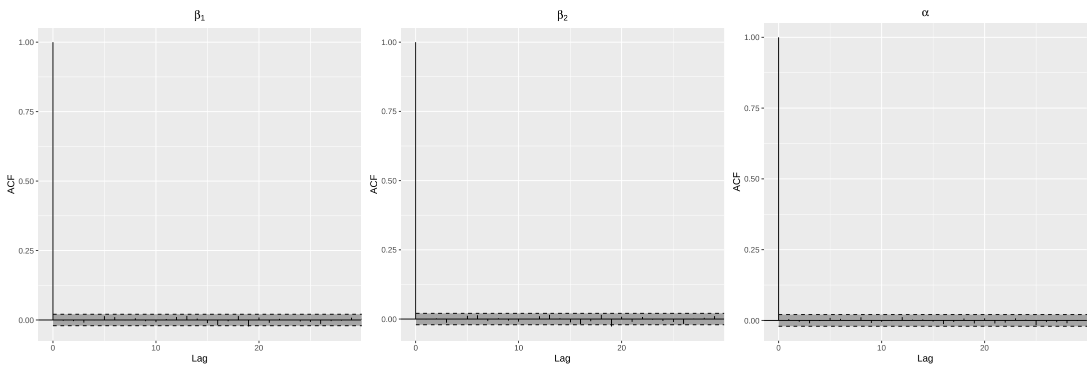
Figure 17 – Simulation Study. Relative error for $\hat{\beta}_3$ under the GL link ((a)-(c)) and the PN link ((d)-(f)) and different sample sizes.

C2 Further results relative to parameter estimation and model adequacy

In this section additional analysis of the bladder cancer dataset are presented. These analysis include posterior densities histograms, trace and autocorrelation plots of the parameters for the PN link with the PC prior, prediction results using the traditional links and residual plots of the selected model.

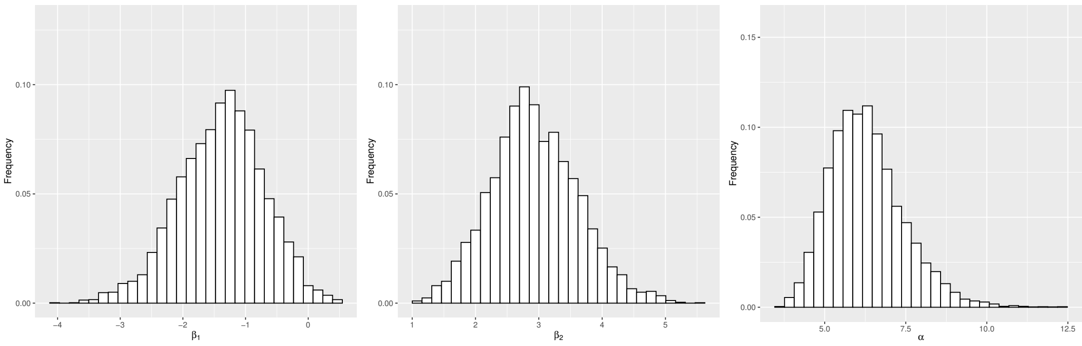


(a) (b) (c)



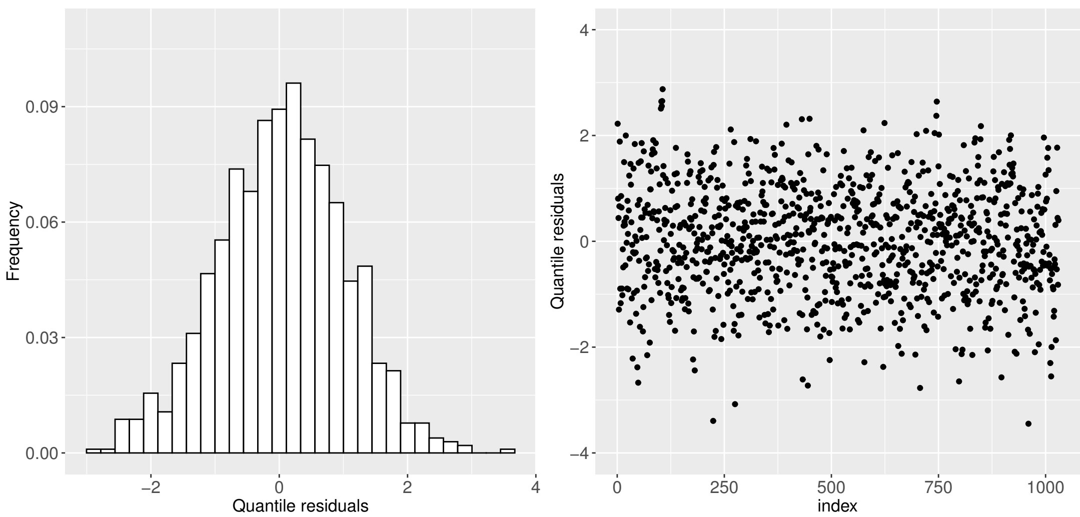
(d) (e) (f)

Figure 18 – Bladder cancer data. Trace and ACF plots for the model parameters considering the PN link with the PC prior.



(a) (b) (c)

Figure 19 – Bladder cancer data. Posterior density for the fixed effects and skewness parameters considering the PN link with the PC prior.



(a) (b)

Figure 20 – Bladder cancer data. Diagnostic plots for the PN link under the PC prior.