Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1–27

Design and Structure Dependent Priors for Scale Parameters in Latent Gaussian Models

Aldo Gardini*, Fedele Greco†, and Carlo Trivisano‡

Abstract. Bayesian inference in latent Gaussian models necessitates the specification of prior distributions for scale parameters, which govern the behavior of model components. This task is particularly delicate and many contributions in the literature are devoted to the topic. We show that the scale parameter plays a crucial role in determining the prior variability of the model components, which is influenced by factors such as correlation structure, design matrices, and potential linear constraints. This intricate relationship adds complexity, making it difficult to interpret and compare priors across diverse applications. To tackle this challenge, we propose a novel approach for prior specification based on the theory of distribution of quadratic forms. Our strategy involves the use of design and structure-dependent (DSD) priors, which ensure a consistent interpretation across diverse applications. By introducing a single parameter that governs the prior variability of the linear predictor, we simplify the process of prior specification, making it more manageable and interpretable. We derive analytical expressions for DSD priors on scale parameters and establish conditions that guarantee their existence. To demonstrate the efficacy of our proposed prior elicitation strategy, we conduct a simulation study, examining the sampling properties of the estimators. Additionally, we explore several real data applications to investigate prior sensitivity and the allocation of explained variance among model components.

Keywords: Gaussian Markov random fields, hypergeometric functions, integral equations, prior elicitation, quadratic forms.

MSC2020 subject classifications: 62F15, 62J12.

1 Introduction

Latent Gaussian Models (LGMs) are a subclass of Generalized Linear Mixed Models where the expected value of a response variable y is connected to a linear predictor η via a link function $g(\eta)$. The linear predictor

$$\boldsymbol{\eta} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta} + \sum_{j=1}^{Q} \mathbf{Z}_j \boldsymbol{\gamma}_j \tag{1}$$

is constituted by a priori independent additive components distributed as Gaussian random variables conditionally on model hyperparameters. The design matrix $\mathbf{X} \in \mathbb{R}^{n \times P}$ is associated to fixed effects $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_P)^\top$ and β_0 is an overall intercept. Random components are expressed as the product of a random effect design matrix $\mathbf{Z}_j \in \mathbb{R}^{n \times m_j}$, with

*Department of Statistical Sciences, University of Bologna, Italy, aldo.gardini@unibo.it
†Department of Statistical Sciences, University of Bologna, Italy
‡Department of Statistical Sciences, University of Bologna, Italy

 https://doi.org/10.1214/24-BA1454

	Intercept	Fixed effects	Random effects
Coefficients	β_0	$\beta_i,\ i = 1, \dots, P$	$\boldsymbol{\gamma}_j,\ j = 1, \dots, Q$
Design	$\mathbf{1}$	$\mathbf{x}_i$	$\mathbf{Z}_j$
Structure	1	1	$\mathbf{K}_{\gamma_j}$
Scaler	10^6	$\sigma^2_{\beta_i}$	$\sigma^2_{\gamma_j}$

Table 1: Adopted notation for components of an LGM.

$m_j \leq n$ and a random vector $\boldsymbol{\gamma}_j \in \mathbb{R}^{m_j}$, $j = 1, \dots, Q$, which in LGMs follows a Gaussian distribution. This setting allows us to encompass models for grouped data, where $\mathbf{Z}_j$ is built as a selection matrix, as well as low-rank models and non-parametric regression, where $\mathbf{Z}_j$ is built as a basis matrix. When the random effect is observation-specific, $\mathbf{Z}_j$ corresponds to the identity matrix $\mathbf{I}_n$. In what follows, we denote observation-specific random effects as the n-dimensional vector $\boldsymbol{\psi}_j = \mathbf{Z}_j\boldsymbol{\gamma}_j$, $j = 1, \dots, Q$. Without loss of generality, we consider both covariates and random effects $\boldsymbol{\psi}_j$ to be centered: this is strongly advised when implementing MCMC algorithms (Gelfand et al., 1995) and constitutes a natural choice for the development of the prior specification strategy proposed in this paper.

We consider priors on $\boldsymbol{\gamma}_j$ with a *fixed* structure matrix $\mathbf{K}_{\gamma_j}$ reflecting the modeler's prior beliefs on the dependence relationships characterizing the j-th random effect. Table 1 summarises the whole model architecture of the LGMs covered in this paper. All model parameters follow a Gaussian distribution:

$$\beta_p|\sigma^2_{\beta_p} \sim \mathcal{N}_1\big(0, \sigma^2_{\beta_p}\big),\ p = 1, \dots, P, \qquad \boldsymbol{\gamma}_j|\sigma^2_{\gamma_j} \sim \mathcal{N}_{m_j}\big(\mathbf{0}, \sigma^2_{\gamma_j}\mathbf{K}^{-1}_{\gamma_j}\big),\ j = 1, \dots, Q,$$

where $\mathbf{K}_{\gamma_j}$ denotes the precision matrix, as we adopt the mean-covariance parameterization of the multivariate Gaussian distribution. The choice of prior distributions for $\boldsymbol{\sigma}^2 = (\sigma^2_{\beta_1}, \dots, \sigma^2_{\beta_P}, \sigma^2_{\gamma_1}, \dots, \sigma^2_{\gamma_Q})$ completes model specification. Such hyperparameters act as mere scalers, governing the contribution of each model component to the total prior variability of the linear predictor. As a measure of such variability, we adopt the sampling variance

$$V_\eta = \frac{1}{n-1}\sum_{i=1}^{n}(\eta_i - \bar{\eta})^2 = \frac{\boldsymbol{\eta}^\top\mathbf{M}\boldsymbol{\eta}}{n-1}, \tag{2}$$

where $\bar{\eta} = n^{-1}\sum_i^n \eta_i$ and $\mathbf{M} = [\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^T]$ is the centering matrix. We deem this choice sensible, recognizing that it is just one among several choices that can be adopted to relate the prior on $\boldsymbol{\sigma}^2$ to model components. Among the others, Klein and Kneib (2016) adopt the absolute value of smooth functions in the context of additive distributional regression, while Wakefield (2007) focus on the range of the relative risks in disease mapping. The merit of sampling variance lies in its ease of interpretation and its ability to establish connections with well-known quantities in the realm of statistical modeling. For instance, in the case of a linear regression model, it corresponds to the regression deviance (divided by $n-1$). Additionally, we believe that considering this quantity for specifying priors on scale parameters allows connections with the insights presented in Gelman et al. (2017), which advocate that the complexities of modern statistical applications require "that we think *generatively* by considering the measurements consistent

with a given prior". In this spirit, we specify prior on scalers $\boldsymbol{\sigma}^2$ by controlling that the induced sampling variance of the linear predictor is appropriate for the phenomenon under study. In other words, priors are defined to ensure that the variability of data generated by the model remains plausible within the context of the phenomenon being studied. In the framework of linear regression with Gaussian likelihood, this can be accomplished by controlling that the prior distribution of V_η is coherent with prior knowledge/expectation about the variability of the phenomenon. As a default strategy that can be customary for practitioners, we suggest using the observed sample variance of the response variable as a reference value in the specification of the prior: more precisely, the user has the ability to specify the prior on scalers so that the data variance matches a chosen quantile of the prior distribution on V_η. We extend this procedure to other GLMs by using the idea of pseudo-variance (Gelman et al., 2013).

Due to the a priori independence between model components, (2) arises as the sum of the prior variability associated with fixed and random effects. The prior specification strategy developed in this paper begins by studying the effect of design matrices, structure matrices, and possible linear constraints in propagating the variability induced by the prior on $\boldsymbol{\sigma}^2$ through the linear predictor. We demonstrate that dependence of design and structure on the application at hand hampers the interpretability of prior specification and that the same prior on a scaler has different implications on (2) because of differences in the model architecture. Our effort is aimed at taking account of such dependence in order to obtain a class of priors that have the same interpretation across different applications. For this reason, we dub the proposed priors as Design and Structure Dependent (DSD) priors: the prior distribution on the sampling variance of each model component, when marginalized with respect to the scaler, will not depend on structure and design. More in detail, resorting on integral equations, we retrieve prior distributions for the scales that always imply the same marginal variance, namely the one of the simple i.i.d. model, regardless of the structure and design associated with the effect. Note that β_0 does not contribute to V_η, and is largely informed by the data. Therefore, as reported in Table 1, we assume for it a non-informative Gaussian prior.

The theoretical developments rely on the theory of Quadratic Forms (QFs) in Gaussian variables, requiring specific computational tools (Gardini et al., 2022). Starting from a Beta distribution of the 2$^{\text{nd}}$ kind as a base prior, we reduce the prior elicitation problem to the specification of one parameter that governs the a priori sampling variance of the model components, which is strictly tied to the degree of smoothness in several models. One of the merits of such a parameter is to allow for intuitive sensitivity analysis, following the same rationale across models that differ with respect to structure and design.

The rest of the paper is organized as follows. Section 2 offers a brief review of the approaches suggested in the literature on priors for scale parameters in LGMs that are relevant to the theory developed in subsequent sections. In Section 3, we introduce our novel prior specification strategy on scalers of the random effects distributions; the prior specification for the fixed effects can be tackled as a special case. A simulation study and real data applications are shown in Sections 4 and 5 respectively, while concluding remarks are offered in Section 6.

2 Hyperpriors on Scale Parameters

As pointed out by Gelman (2006) and Polson and Scott (2012), selecting priors for scale parameters of hierarchical models requires particular care. Furthermore, if the scale parameter is set to 0, the model reduces to the one without the random effect (labeled as base model in the literature): Gelman (2006) advises priors with non-null density function at 0. The idea of base model is also a cornerstone of the research line on Penalizing-Complexity (PC) priors opened by Simpson et al. (2017). PC priors necessitate the specification of an appropriate base model that represents a simplified version of the (more complex) posited model. The prior is then defined based on the distance between the base and the posited model and it is characterized by a non-null density at zero, favoring shrinkage towards the base model.

Besides the distributional assumption presented in Section 2.1, hyperpriors also need care in calibration and, possibly, elicitation of prior knowledge in the model. For these tasks, assumptions on the model components and the available auxiliary information play a crucial role that must be taken into account, as will be discussed in Section 2.2. Section 2.3 completes the literature review on the topic, listing alternative approaches to the one pursued in the paper.

2.1 Prior Distribution: the Beta Distribution of the 2nd Kind

This section introduces the probability distribution that plays a central role in the prior specification strategy proposed in the paper. As noted by Perez et al. (2017), most of the priors proposed for scale parameters are embedded within the Beta distribution of the 2nd kind, henceforth $\sigma^2 \sim \text{B2}(b, p, q)$, which is a special case of the Generalized Beta Prime distribution. The B2 distribution is ruled by a scale parameter b and two shape parameters p and q, the density function is:

$$f^{\text{B2}}_{\sigma^2}(s) = \mathcal{K}_{\text{B2}}\, s^{-q-1}\left(1+\frac{b}{s}\right)^{-p-q}, \tag{3}$$

where $\mathcal{K}_{\text{B2}} = \frac{b^q}{B(p,q)}$ is the normalizing constant and $B(\cdot,\cdot)$ is the beta function. Parameters p and q control the behavior of the distribution tails: p determines the behavior of the density at 0 (divergent for $p < 1$, finite for $p = 1$ and 0 for $p > 1$), and q regulates the degree of the polynomial decay of the right tail (the higher q, the lighter the tail). In addition, a B2 prior on σ^2 with $p = 0.5$ implies a finite non-null density at 0 for σ.

The interpretability of the parameters and the flexibility they deliver to the distribution represent arguments in favor of the choice of a B2 prior on σ^2. In addition, the distribution includes several popular priors as special cases. The widely used Half-t prior on σ with d degrees of freedom and scale c coincides with a $\text{B2}(c^2 d, 1/2, d/2)$ prior for σ^2. Gamma and Inverse Gamma distributions are obtained as limiting cases: $\sigma^2 \sim \text{Gamma}(p, c) = \lim_{q\to+\infty} \text{B2}(q/c, p, q)$ and $\sigma^2 \sim \text{Inv-Gamma}(q, c) = \lim_{p\to+\infty} \text{B2}(p/c, p, q)$.

2.2 The Impact of Structure and Design

The strategy for prior specification that we propose is designed to acknowledge the influence that both the design matrix and the structure of a model component can exert on its prior variability. In what follows, we provide an overview of significant contributions that focus on the impact of model architecture on prior configuration.

A first element of criticism emerges whenever a non-diagonal precision matrix is assumed for the random effects distribution, as in this case interpretation of the prior can hardly be disentangled by the structure matrix (Section 3.1 is devoted to the formalization of such dependence). As an example, in the framework of disease mapping, Bernardinelli et al. (1995) noted that the interpretation of a scale parameter of a spatial random effect is conditional with respect to the neighborhood structure and, therefore, different priors should be assigned to the structured and unstructured components. More generally, this consideration can be extended to each case in which an intrinsic Gaussian Markov random field (IGMRF, Rue and Held, 2005) prior is assumed for a random effect, leading to a rank deficient structure matrix $\mathbf{K}_{\gamma_j}$. Indeed, IGMRFs are in general heteroscedastic, being the variance dependent on the structure. For this reason, Sørbye and Rue (2014) suggest to specify a scaled hyperprior for $\sigma^2_{\gamma_j}$, accounting for the structure. In particular, they specify a prior on $\sigma^2_{\gamma_j} s^2_{ref,j}$, where $s_{ref,j}$ is the reference standard deviation for the considered IGMRF. Such scaling value is set equal to the geometric mean of the non-null eigenvalues of the generalized inverse $\mathbf{K}^-_{\gamma_j}$. The procedure is implemented in the popular `R-INLA` software (Rue et al., 2009) and has been exploited by Riebler et al. (2016) in the context of the popular BYM model (Besag et al., 1991).

Furthermore, if $\boldsymbol{\psi}_j = \mathbf{Z}_j \boldsymbol{\gamma}_j$, with $\mathbf{Z}_j \neq \mathbf{I}_n$, the design matrix also contributes to the prior variability. For this reason, Klein and Kneib (2016) remark that it is not possible to elicit the prior information only considering $\mathbf{K}_{\gamma_j}$ since the sampling variance of the whole vector of random effects $\boldsymbol{\psi}_j$ would be ignored. The authors focused on the P-splines regression framework and their strategy to elicit a prior on $\sigma^2_{\gamma_j}$ relies on a probabilistic statement about the range of variation of $\boldsymbol{\psi}_j$, as a measure of its dispersion. The scale parameter of the prior is obtained after marginalizing the distribution of the range over $\sigma^2_{\gamma_j}$. We remark that the chosen measure accounts both for the structure of the random effect $\mathbf{K}_{\gamma_j}$ and the covariate patterns involved in $\mathbf{Z}_j$. This represents an interesting contribution with some strict connections with the strategy we propose. Indeed, Klein and Kneib (2016) suggest to set the prior variability of model effects by relying on the whole vector $\boldsymbol{\psi}_j$, instead of the sole coefficients $\boldsymbol{\gamma}_j$.

2.3 Other Approaches: Splitting the Total Variability

An alternative approach to the problem of specifying the priors and hyperpriors in hierarchical models is based on the idea of splitting the overall model variance among the different components. This can be thought as an extension of the prior on the coefficient of determination R^2 for linear models proposed by Gelman and Hill (2006). While our approach is not directly focused on splitting the total variability, we found these ideas

relevant to our developments as they relate scale parameters to the contribution of each model component at the linear predictor level.

We mention two recent interesting proposals within this framework. Narrowing the field to the linear model case, Zhang et al. (2020) derived a shrinkage prior for the regression coefficient, specifying a marginal prior on the R^2 coefficient. An interesting feature of such a strategy is that the design is automatically integrated out, deducing the induced priors on the vector of coefficients that are in practice dependent on the design. This framework has been extended in several directions: Yanchenko et al. (2021) face the problem of generalizing this procedure to Generalized Linear Mixed Models, Aguilar and Bürkner (2023) focus on Linear Multilevel Models, and Yanchenko et al. (2023) deal with the case of spatial regression. With a similar spirit, the work by Fuglstad et al. (2020) proposes to specify a prior distribution for the total variability and then split it among the distinct model components, accounting for their structure, putting a PC prior on the splitting nodes.

3 Design and Structure Dependent Priors

Design and structure dependent priors are introduced by focusing on the random effect $\boldsymbol{\psi}_j = \mathbf{Z}_j\boldsymbol{\gamma}_j$. The specification of the prior on the scale parameter is based on its effect on the sampling variance of $\boldsymbol{\psi}_j$ that, conditionally on $\sigma^2_{\gamma_j}$, is defined as the random variable:

$$V_{\psi_j}|\sigma^2_{\gamma_j} = \frac{\sigma^2_{\gamma_j}}{n-1}\sum_{i=1}^{n}(\psi_{ij} - \bar{\psi}_j)^2 = \sigma^2_{\gamma_j}\frac{\boldsymbol{\psi}_j^\top\mathbf{M}\boldsymbol{\psi}_j}{n-1} = \sigma^2_{\gamma_j}\frac{\boldsymbol{\gamma}_j^\top\mathbf{Z}_j^\top\mathbf{M}\mathbf{Z}_j\boldsymbol{\gamma}_j}{n-1}. \tag{4}$$

This is a QF in Gaussian variables, whose distribution depends on both the design matrix $\mathbf{Z}_j$ and the structure matrix $\mathbf{K}_{\gamma_j}$ (for further details, see Section S2 in the online Supplementary Material, Gardini et al., 2024a). Note that the structure matrix induced on $\boldsymbol{\psi}_j$ is the generalised inverse of $\mathbf{Z}_j\mathbf{K}^{-1}_{\gamma_j}\mathbf{Z}_j^\top$, i.e. $\mathbf{K}_{\psi_j} = (\mathbf{Z}_j\mathbf{K}^{-1}_{\gamma_j}\mathbf{Z}_j^\top)^-$.

The prior distribution on $\sigma^2_{\gamma_j}$ is specified in order to control the marginal distribution of V_{ψ_j}. In this way, we make the prior on $\sigma^2_{\gamma_j}$ dependent on the design and the structure, with the aim of making the marginal prior on V_{ψ_j} independent on both of them. The conditional and marginal distributions of V_{ψ_j} are investigated in the following section.

3.1 Conditional and Marginal Priors on V_{ψ_j}

The distribution of V_{ψ_j} conditional on $\sigma^2_{\gamma_j}$ can be expressed as a linear combination of independent chi-square random variables (Box, 1954; Ruben, 1962),

$$V_{\psi_j}|\sigma^2_{\gamma_j} \sim \frac{\sigma^2_{\gamma_j}}{n-1}\sum_{k=1}^{n}\lambda_{kj}X_k; \quad X_k \overset{ind}{\sim} \chi^2_1, \quad k = 1,\dots,n, \tag{5}$$

where $\boldsymbol{\lambda}_j = (\lambda_{1j}, \ldots, \lambda_{nj})^\top = \text{eig}(\mathbf{M}\mathbf{K}_{\psi_j}^{-1})$. Expectation and variance are:

$$\mathbb{E}\big[V_{\psi_j}|\sigma^2_{\gamma_j}\big] = \frac{\sigma^2_{\gamma_j}}{n-1}\sum_{k=1}^{n}\lambda_{kj} \text{ and } \mathbb{V}\big[V_{\psi_j}|\sigma^2_{\gamma_j}\big] = \left(\frac{\sigma^2_{\gamma_j}}{n-1}\right)^2\sum_{k=1}^{n}2\lambda^2_{kj}.$$

We remark that at least one of the eigenvalues is null because of the multiplication by the rank-deficient matrix $\mathbf{M}$. Equation (5) reveals that the dependence of the distribution of $V_{\psi_j}|\sigma^2_{\gamma_j}$ on the structure matrix $\mathbf{K}_{\psi_j} = (\mathbf{Z}_j\mathbf{K}^{-1}_{\gamma_j}\mathbf{Z}_j^\top)^-$ is captured by the eigenvalues $\boldsymbol{\lambda}_j$: different structure and design matrices lead to different conditional moments of V_{ψ_j}. For this reason, a naive prior specification that chooses the same prior distribution for each $\sigma^2_{\gamma_j}$ would imply different marginal distributions of the sampling variances V_{ψ_j}, i.e. different contributions of each random component to the a priori sampling variance of the linear predictor.

Following analogous arguments, Sørbye and Rue (2014) proposed to scale the structure matrices by the geometric mean of their eigenvalues: a similar approach would consist in dividing $\mathbf{K}_{\psi_j}$ by $\frac{\sum_{k=1}^{n}\lambda_{kj}}{n-1}$. This removes dependence of $\mathbb{E}[V_{\psi_j}|\sigma^2_{\gamma_j}]$ on $\mathbf{K}_{\psi_j}$, obtaining that $\mathbb{E}[V_{\psi_j}|\sigma^2_{\gamma_j}] = \sigma^2_{\gamma_j}$, $\forall j$. Nonetheless, scaling $\mathbf{K}_{\psi_j}$ by any constant does not completely remove the dependence of the marginal density $f_{V_{\psi_j}}$ on $\mathbf{K}_{\psi_j}$. In other words, if the same prior density f_{σ^2} is selected for each $\sigma^2_{\gamma_j}$, the marginal distributions

$$f_{V_{\psi_j}}(v) = \int_0^{+\infty} f_{V_{\psi_j}|\sigma^2_{\gamma_j}}(v|s)f_{\sigma^2}(s)\mathrm{d}s, \quad j = 1, \ldots, Q, \tag{6}$$

are all different because $f_{V_{\psi_j}|\sigma^2_{\gamma_j}}(\cdot) \neq f_{V_{\psi_l}|\sigma^2_{\gamma_l}}(\cdot)$ for each $j \neq l$. In summary:

$$f_{\sigma^2_{\gamma_j}} \equiv f_{\sigma^2},\ \forall j \implies f_{V_{\psi_j}} \not\equiv f_{V_{\psi_l}},\ \forall l \neq j;$$

confirming that equal priors on scale parameters lead to different priors on random effects sampling variances, even after scaling by a constant.

We stress that assigning different prior importance to each random component is perfectly sensible if it reflects the modeler's prior beliefs: this can be beneficial and even desirable in Bayesian applications. Several efforts have been made in the literature to build intuitive and flexible prior specification strategies that allow managing the contributions of each random component to the total variability, a recent prominent example being Fuglstad et al. (2020). The approach proposed in what follows can be generalized to settings where prior unbalance between component-wise contribution is desired: this will be the object of future extensions of the present paper.

In Section 3.2, the DSD prior for the case of full-rank precision matrices $\mathbf{K}_{\gamma_j} \succ 0$ is derived, whereas the important case of semi-positive definite matrices $\mathbf{K}_{\gamma_j}$, which includes the class of IGMRF priors, is considered in Section 3.3. Section 3.4 is devoted to prior elicitation.

3.2 Derivation of the DSD Prior

As a starting point, we focus on the effect of the structure matrix by considering a single random effect $\boldsymbol{\psi}|\sigma^2 \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{K}_\psi^{-1})$, with $\mathbf{Z} = \mathbf{I}_n$ and $\mathbf{K}_\psi = \mathbf{K}_\gamma \succ 0$. The structure

matrix $\mathbf{K}_\psi$ encodes the conditional relationships between elements of $\boldsymbol{\psi}$ and it depends on the application at hand.

A notable special case, that plays a prominent role in the proposed prior specification strategy, is represented by i.i.d. random effects obtained by setting $\mathbf{K}_\psi = \mathbf{I}_n$. This is an ideal benchmark for managing prior specification on scale parameters because of its simplicity and interpretability: in the i.i.d. model σ^2 corresponds to both the expected sampling variance of $\boldsymbol{\psi}$ and to the variance of each component ψ_i, independently on n. The sampling variance of an i.i.d. random effect, denoted as $\mathcal{V}$, is distributed as a scaled chi-square with $n-1$ degrees of freedom conditionally on σ^2. Equivalently, posing $\alpha = \frac{n-1}{2}$ and $\beta = \frac{n-1}{2}$, one obtains $\mathcal{V}|\sigma^2 \sim \text{Gamma}(\alpha, \frac{\beta}{\sigma^2})$, with $\mathbb{E}[\mathcal{V}|\sigma^2] = \mathbb{V}[\psi_i|\sigma^2] = \sigma^2$. Given the generality of the distribution, we opt for a B2 prior (introduced in Section 2.1) on σ^2 in the i.i.d. case. Consequently, the marginal distribution of $\mathcal{V}$ is a mixture of a gamma and a B2 distributions whose density function is given in Proposition 3.1. Since this marginal density can be expressed in terms of the ${}_2F_0$ hypergeometric function (Olver et al., 2010), we dub it ${}_2\mathcal{F}_0$-distribution.

Proposition 3.1 (${}_2\mathcal{F}_0$-distribution)**.** *Let X and Y be two random variables such that*

$$X|y \sim \text{Gamma}\left(\alpha, \frac{\beta}{y}\right), \quad Y \sim \text{B2}(b, p, q).$$

Then, $X \sim {}_2\mathcal{F}_0(\alpha, \beta/b, p, q)$ where α, p and q are shape parameters and the scale parameter β/b is the ratio of the scale parameters of the mixed distributions. The density of X is:

$$f_X(x) = \left(\frac{b}{\beta}\right)^q \frac{\Gamma(1+\alpha-p)}{\Gamma(p)B(\alpha,q)} x^{-q-1}\ {}_2F_0\left(\alpha+q, p+q; -; -\frac{b}{x\beta}\right). \tag{7}$$

Proof. See Section S3 of the online Supplementary Material (Gardini et al., 2024a). □

Thus, for the i.i.d. model, the marginal distribution of $\mathcal{V}$ is

$$\mathcal{V} \sim {}_2\mathcal{F}_0(\alpha, \beta/b, p, q). \tag{8}$$

However, if $\mathbf{K}_\psi \neq \mathbf{I}_n$, the sampling variance of the random effect conditioned on σ^2 is a linear combination of chi-squared random variables (see (5)). Hence, a B2(b, p, q) prior on σ^2 would result in a different marginal distribution of V_ψ, which depends on the application at hand: we argue that this dependence is undesirable since it impedes coherence of prior statements among different models. To remove such dependence, we aim to obtain a DSD prior on σ^2, denoted as $f_{\sigma^2}^{DSD}$, that, when mixed with the distribution of $V_\psi|\sigma^2$, delivers the density of the benchmark ${}_2\mathcal{F}_0(\alpha, \beta/b, p, q)$ distribution. Technically, we seek the prior density which solves the integral equation:

$$f_{\mathcal{V}}(v) = \int_0^{+\infty} f_{V|\sigma^2}(v|s) f_{\sigma^2}^{DSD}(s) \mathrm{d}s.$$

The merit of such distribution would be to have the same interpretation, independently on $\mathbf{K}_\psi$, in terms of marginal variance, which will always coincide with that of the i.i.d. model.

To obtain a mathematically and numerically tractable problem, we develop DSD priors working on the following approximation of $V|\sigma^2$:

$$V|\sigma^2 \overset{a}{\sim} \text{Gamma}\left(\tilde{\alpha}, \frac{\tilde{\beta}}{\sigma^2}\right), \tag{9}$$

where $\overset{a}{\sim}$ indicates a random variable *approximately distributed as*. This approximation has been proposed by Box (1954) in order to match the first two moments of $V|\sigma^2$, and parameters are

$$\tilde{\alpha} = \frac{(\sum_{i=1}^n \lambda_i)^2}{2\sum_{i=1}^n \lambda_i^2}, \qquad \tilde{\beta} = \frac{n-1}{2}\frac{\sum_{i=1}^n \lambda_i}{\sum_{i=1}^n \lambda_i^2}. \tag{10}$$

Given that $V|\sigma^2$ approximately follows a Gamma distribution, Proposition 3.1 implies that, if $\sigma^2 \sim \text{B2}(b, p, q)$, the approximate marginal distribution of V is

$$V \overset{a}{\sim} {}_2\mathcal{F}_0(\tilde{\alpha}, \tilde{\beta}/b, p, q). \tag{11}$$

Comparison between (8) and (11) clearly highlights that the difference between these marginal distributions is due to the eigenvalues of $\mathbf{M}\mathbf{K}_\psi^{-1}$. The following theorem constitutes the main result of the paper, stating the density of the DSD prior, i.e. a prior that delivers the marginal density (8) independently on $\mathbf{K}_\psi$. The prior density turns out to be expressed in terms of the ${}_2F_1$ hypergeometric function (Olver et al., 2010).

Theorem 3.1 (DSD prior)**.** *Let $\boldsymbol{\psi}|\sigma^2 \sim \mathcal{N}_n(\mathbf{0}, \sigma^2\mathbf{K}_\psi^{-1})$. The DSD prior $f_{\sigma^2}^{DSD}$, i.e. the prior that solves the integral equation*

$$\int_0^{+\infty} f_{V|\sigma^2}(v|s) f_{\sigma^2}^{DSD}(s)\mathrm{d}s = f_{\mathcal{V}}(v), \tag{12}$$

where $\mathcal{V} \sim {}_2\mathcal{F}_0(\alpha, \beta/b, p, q)$ and $V|\sigma^2 \overset{a}{\sim} Gamma(\tilde{\alpha}, \frac{\tilde{\beta}}{\sigma^2})$, has density

$$f_{\sigma^2}^{DSD}(s) = \mathcal{K}_{DSD}\, s^{-q-1}\, {}_2F_1\left(q+\alpha, q+p; q+\tilde{\alpha}; -\frac{b\tilde{\beta}}{s\beta}\right), \tag{13}$$

with

$$\mathcal{K}_{DSD} = \left(\frac{b\tilde{\beta}}{\beta}\right)^q \frac{1}{B(p,q)}\frac{\Gamma(\tilde{\alpha})}{\Gamma(q+\tilde{\alpha})}\frac{\Gamma(q+\alpha)}{\Gamma(\alpha)},$$

provided that $p \leq \tilde{\alpha}$.

Proof. See Section S4 in the online Supplementary Material (Gardini et al., 2024a). □

Firstly, it is worth noting that the existence of DSD priors is guaranteed by setting $p \leq 1/2$, regardless of the model structure and design. This can be observed from (10), which implies that $\tilde{\alpha} \geq 1/2$ since $(\sum_{i=1}^n \lambda_i)^2 / \sum_{i=1}^n \lambda_i^2 \geq 1$. On the other hand, recalling from Section 2.1 that the popular Inverse Gamma prior on σ^2 is obtained letting $p \to +\infty$, DSD priors are not defined when an Inverse Gamma is specified as a base prior.

A second comment regards the impact of the approximation introduced in (9). Recalling that the true distribution of $V|\sigma^2$ is a linear combination of chi-square variables with positive weights, it can be shown that, by applying Proposition 3.1, the marginal distribution of V is a weighted sum of ${}_2\mathcal{F}_0$-distributions. However, to get the results stated in Theorem 3.1, namely the DSD prior density after proving that it represents a valid probability distribution, the Mellin transform of $V|\sigma^2$ is required. As shown in Gardini et al. (2022), the Mellin transform of a positive definite quadratic form can be expressed as an infinite sum of ratios of gamma functions: its use in solving the integral equation (12) would lead to an integral without closed form solution, up to our knowledge. The formal results behind these considerations are shown in Section S5 of the Supplementary Material (Gardini et al., 2024a). Conversely, approximating the quadratic form distributions through a single Gamma distribution with the first two moments matched, as proposed by Box (1954), represents an appealing trade-off between keeping the problem tractable and preserving key features of the approximated distribution, as also shown in Figure S1 in the Supplementary Material (Gardini et al., 2024a).

Comparison between (3) and (13) clearly shows how DSD priors modify the base B2 distribution. Indeed, the kernels of these distributions differ with respect to the last factor: they both lay in the interval $(0, 1)$ and are increasing functions of s. The couples (α, β) and $(\widetilde{\alpha}, \widetilde{\beta})$ determine the growth rate of such factors leading to different shapes of DSD prior with respect to the base B2 prior: an example of such modification is reported in Figure 1. As a notable special case of Theorem 3.1, due to the properties of the ${}_2F_1$ function, if $V = \mathcal{V}$, i.e. $\tilde{\alpha} = \alpha$ and $\tilde{\beta} = \beta$, the prior reduces to $\sigma^2 \sim B2(b, p, q)$. Furthermore, if $p = \tilde{\alpha}$ the DSD prior is $\sigma^2 \sim \text{B2}(\frac{b\tilde{\beta}}{\beta}, \alpha, q)$.

The dependence of the prior on $\mathbf{K}_\psi$ is captured by $\widetilde{\alpha}$ and $\widetilde{\beta}$ as defined in (10): all the generalizations proposed in what follows are based on the fact that, to obtain the DSD prior, one must consider the appropriate eigenvalues vector, i.e. those that express the weights of the associated QF $V_\psi|\sigma^2$.

Thus, taking account of the effect of the design is quite a straightforward task. Indeed, when $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\gamma}$ with $\mathbf{Z} \in \mathbb{R}^{n\times m}$ and $\boldsymbol{\gamma}|\sigma^2 \sim \mathcal{N}_m(\mathbf{0}, \sigma^2\mathbf{K}_\gamma^{-1})$, with $\mathbf{K}_\gamma \succ 0$, the appropriate eigenvalues are those of the matrix $\mathbf{MZK}_\gamma^{-1}\mathbf{Z}^\top$ whose non-null elements coincide with non-null eigenvalues of $\mathbf{Z}^\top\mathbf{MZK}_\gamma^{-1}$.

The very special case of $m = 1$ is useful to shed light on the rationale behind DSD priors. Indeed, this corresponds to managing a fixed effect, and, for this reason, we switch the notation to $\boldsymbol{\psi} = \mathbf{x}\beta$ where $\mathbf{x} \in \mathbb{R}^{n\times 1}$ and $\beta|\sigma^2 \sim \mathcal{N}(0, \sigma^2)$. The sampling variance

$$V_{\boldsymbol{\psi}} = \frac{\boldsymbol{\psi}^\top\mathbf{M}\boldsymbol{\psi}}{n-1} = \frac{\mathbf{x}^\top\mathbf{M}\mathbf{x}}{n-1}\beta^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}\beta^2 = \beta^2 s_x^2,$$

has conditional distribution $V_{\boldsymbol{\psi}}|\sigma^2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2\sigma^2 s_x^2})$, arising as the square of a zero-mean Gaussian distribution with variance σ^2 and multiplied by the only non-null weight of the QF, i.e. s_x^2. If $p = \frac{1}{2}$, as will be justified in Section 3.4, the DSD prior on σ^2 turns out to be $\sigma^2 \sim \text{B2}(\frac{b}{s_x^2}, \alpha, q)$. As a consequence, in a model with P fixed effects, the same prior distribution on all standardized coefficients arises.

Lastly, focusing on computational aspects, we provide the `DSDprior` package in the Supplementary Material (Gardini et al., 2024b). This `R` package enables the management of DSD priors and implements standard MCMC algorithms for estimating LGM models with fixed structure matrices. The package allows for block sampling from the full conditional distributions of $(\boldsymbol{\gamma}_j, \sigma^2_{\gamma_j})$, $j = 1, \ldots, Q$, using a Metropolis step, as suggested in Knorr-Held and Rue (2002). Details on the implementation of the MCMC algorithm are given in Section S6 of the Supplementary Material (Gardini et al., 2024a), while Section S7.2 gives some details on convergence and efficiency. The computational complexity associated with DSD priors arises from the evaluation of the ${}_2F_1$ hypergeometric function to obtain acceptance probabilities for the Metropolis steps. Some details on computational times are given in Section 5.

3.3 The Case of Intrinsic Priors

IGMRF priors, which are commonly used in spatial and spatio-temporal modeling of areal data as well as in low-rank models, are improper priors with sparse rank-deficient precision matrices. Let $\text{Rank}(\mathbf{K}_\psi) = n - \kappa$: the rank-deficiency κ is defined as the order of the IGMRF by Rue and Held (2005). Again, we start by assuming $\mathbf{Z} = \mathbf{I}_n$. The developments in what follows are based on the spectral decomposition $\mathbf{K}_\psi = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \mathbf{U}_+\boldsymbol{\Lambda}_+\mathbf{U}_+^\top$, where $\boldsymbol{\Lambda}_+ \in \mathbb{R}^{(n-\kappa)\times(n-\kappa)}$ is a diagonal matrix with diagonal entries corresponding to the non-null eigenvalues of $\mathbf{K}_\psi$ and $\mathbf{U}_+$ spans the column space of $\mathbf{K}_\psi$; in addition, $\mathbf{U}_0$ spans the null space.

From a probabilistic viewpoint, an IGMRF of order κ embeds a proper distribution on the $(n-\kappa)$-dimensional column space of $\mathbf{K}_\psi$ describing deviations from the κ-dimensional null space, i.e. the implicit systematic part of the model. This important feature of IGMRF priors is highlighted in Rue and Held (2005), Section 3.4.1, where $\boldsymbol{\psi}$ is decomposed as:

$$\boldsymbol{\psi} = \text{trend}(\boldsymbol{\psi}) + \text{residuals}(\boldsymbol{\psi}) = \mathbf{U}_0\mathbf{U}_0^\top\boldsymbol{\psi} + \mathbf{U}_+\mathbf{U}_+^\top\boldsymbol{\psi} = \mathbf{U}_0\boldsymbol{\psi}_0 + \mathbf{U}_+\boldsymbol{\psi}_+, \tag{14}$$

and $\boldsymbol{\psi}_+|\sigma^2 \sim \mathcal{N}_{n-\kappa}(\mathbf{0}, \sigma^2\boldsymbol{\Lambda}_+^{-1})$. This decomposition has also been used in Goicoa et al. (2018) to study the need for linear constraints in spatio-temporal disease mapping and in Klein and Kneib (2016) for obtaining their scale dependent priors.

Identifiability of $\boldsymbol{\psi}$ can be ensured by adopting the linear constraint $\mathbf{U}_0^\top\boldsymbol{\psi} = \mathbf{0}$, as suggested in Schrödle and Held (2011). When implementing LGMs with IGMRF priors, we include in the linear predictor the columns of $\mathbf{U}_0$ that are not linearly dependent on other model components. The prior on $\boldsymbol{\psi}_0$ is specified as in Section 3.2. To give some examples, if a random walk (RW) prior of order 1 is employed, the null space is spanned by the unit vector and hence it is already included in the intercept term, while the null space of a RW of order 2 is spanned by a first-order polynomial, requiring the inclusion of a linear trend.

The sampling variance of a *constrained* IGMRF random effect is

$$V_{\boldsymbol{\psi}|\mathbf{U}_0^\top\boldsymbol{\psi}=\mathbf{0}} = \frac{\boldsymbol{\psi}_+^\top\mathbf{U}_+^\top\mathbf{M}\mathbf{U}_+\boldsymbol{\psi}_+}{n-1}, \tag{15}$$

with $\boldsymbol{\lambda} = \text{eigen}(\mathbf{U}_+^\top \mathbf{M}\mathbf{U}_+ \boldsymbol{\Lambda}_+^{-1}) = \text{eigen}(\mathbf{M}\mathbf{K}_\psi^-)$ being the eigenvalues to be taken into account for computing $\widetilde{\alpha}$ and $\widetilde{\beta}$.

This can be directly extended to the case $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\gamma}$, $\mathbf{Z} \in \mathbb{R}^{n \times m}$, $\boldsymbol{\gamma}|\sigma^2 \sim \mathcal{N}_m(\mathbf{0}, \sigma^2 \mathbf{K}_\gamma^-)$. Given the decomposition $\boldsymbol{\gamma} = \mathbf{U}_{\gamma 0}\boldsymbol{\gamma}_0 + \mathbf{U}_{\gamma +}\boldsymbol{\gamma}_+$ obtained by applying (14) to $\boldsymbol{\gamma}$, one gets $\boldsymbol{\psi} = \mathbf{Z}\mathbf{U}_{\gamma 0}\boldsymbol{\gamma}_0 + \mathbf{Z}\mathbf{U}_{\gamma +}\boldsymbol{\gamma}_+$. Introducing the linear constraint $\mathbf{U}_{\gamma 0}^\top \mathbf{Z}^\top \boldsymbol{\psi} = \mathbf{U}_{\gamma 0}^\top \mathbf{Z}^\top \mathbf{Z}\boldsymbol{\gamma} = \mathbf{0}$, the sampling variance is

$$V_{\boldsymbol{\psi}|\mathbf{U}_{\gamma 0}^\top \mathbf{Z}^\top \boldsymbol{\psi} = \mathbf{0}} = \frac{\boldsymbol{\gamma}_+^\top \mathbf{U}_{\gamma +}^\top \mathbf{Z}^\top \mathbf{M}\mathbf{Z}\mathbf{U}_{\gamma +}\boldsymbol{\gamma}_+}{n-1}, \tag{16}$$

i.e. a QF whose weights are the non-null eigenvalues of the semi-positive definite matrix $\mathbf{U}_{\gamma +}^\top \mathbf{Z}^\top \mathbf{M}\mathbf{Z}\mathbf{U}_{\gamma +}\boldsymbol{\Lambda}_{\gamma +}^{-1}$ that coincide with the non-null eigenvalues of $\mathbf{Z}^\top \mathbf{M}\mathbf{Z}\mathbf{K}_\gamma^-$. The computation of the eigenvalues of such a matrix allows us to completely determine the DSD priors by retrieving $\widetilde{\alpha}$ and $\widetilde{\beta}$.

It is worth noting that linear constraints on random effects envisioned in this section guarantee property of the posterior distribution: indeed, linear constraints address partial improperty of IGMRF priors, by removing the improper prior on the implicit systematic part of the model. For a comprehensive discussion on this topic in the context of structured additive distributional regression, see Klein and Kneib (2016).

3.4 Prior Elicitation

The aim of this section is to deliver a simple and intuitive prior elicitation strategy that can be adopted for every LGM whose architecture falls within Table 1. This is facilitated by DSD priors due to their ability to trace back the prior interpretation within the framework of the i.i.d. model, regardless of the specific LGM being considered. The B2 distribution serves as a base prior to be adapted to the considered model by solving integral equation (12): this guarantees the same marginal distribution of the sampling variances $f_{V_{\psi_j}} \equiv f_{\mathcal{V}}$, $\forall j$. As a starting point, we suggest default values of the shape parameters p and q. As regards the scale parameter b, we find it convenient to set it by considering observed data variability.

To guarantee a finite non-null density at 0, we set $p = 1/2$, leading to the specification of an Half-t distribution on σ, as suggested in Perez et al. (2017). Moreover, this choice preserves the conditions for the existence of DSD priors stated in Theorem 3.1, given that $\tilde{\alpha} > 1/2$ holds for every LGM. We suggest $q = 1.5$ as a default choice: this delivers an Half-t with 3 degrees of freedom on σ. However, we show in subsequent simulations and applications that posterior inference shows small sensitivity to q.

The scale parameter b is the crucial quantity to be specified, because of its impact on the amount of shrinkage/smoothness of the posterior estimates. Linking this parameter to the variability of the linear predictor allows automatic scaling of the prior to the considered application. Such scaling is achieved by means of a probability statement on the variability of the j-th model component, in the same spirit of Wakefield (2007); Klein and Kneib (2016) and Simpson et al. (2017), among others. In particular, we

control the probability π_0 that the marginal sampling variance of a model component $\boldsymbol{\psi}_j$ is lower than a value c:

$$\mathbb{P}[V_{\psi_j} \leq c] = \pi_0, \quad \forall j. \tag{17}$$

There are various strategies and philosophies for determining the value of c. As a practical and readily implementable choice for practitioners, we find it appropriate to calibrate the prior variability of the linear predictor with the variability observed in the sample. Despite the double use of the data, this strategy is currently widespread and forms the foundation of the default prior setting in Gaussian models fitted using the popular `rstanarm` package. (Goodrich et al., 2022). For example, when the likelihood is Gaussian, we suggest to set $c = s_y^2$, where s_y^2 denotes the sample variance of the response. Hence, the data variance is the π_0-th quantile of the marginal distribution of V_{ψ_j}, $\forall j$: the higher π_0 the lower the prior variability, with consequent heavier shrinkage. In the case of non-Gaussian likelihood, we find it appealing to resort to the concept of pseudo-variance (Gelman et al., 2013). Such quantity can be exploited to specify meaningful values of c resorting to the observed sample or available prior knowledge on the variability of the phenomenon. This is not an easy task when the model is not Gaussian and the link is not the identity. For an overview of GLMs pseudo-variances under canonical link functions, see Table 1 in Piironen and Vehtari (2017). Once c and π_0 are fixed, b is retrieved by solving the equation $\mathbb{P}[bV^*_{\psi_j} \leq c] = \pi_0$, where $V^*_{\psi_j} = V_{\psi_j}/b \sim {}_2\mathcal{F}_0(\alpha, \beta, p, q)$, via Monte Carlo simulation. Exploiting the mixture representation of the ${}_2\mathcal{F}_0$-distribution, M replicates from $V^*_{\psi_j}$ are generated for computing the quantile corresponding to probability π_0, denoted with $c^*_{\pi_0}$. Consequently, the value of b is retrieved as $c/c^*_{\pi_0}$.

Summarizing, the base prior B2$(b = f(y, \pi_0), 0.5, 1.5)$ is suggested as a default choice, so that prior specification of a given LGM reduces to set the parameter π_0, while DSD priors filter out the effect of design and structure matrices as well as possible linear constraints that hamper comparability of priors in different models. In Sections 4 and 5 we show that π_0 is a decisive quantity in terms of impact on posterior inference: the study of the variation of posterior inference with respect to π_0 is able to give insights on sensitivity to prior specification.

3.5 Decomposition of the Linear Predictor Sampling Variance

To provide a further illustration of the rationale behind DSD priors, we discuss the a priori decomposition of the linear predictor sampling variance marginally with respect to scalers $\boldsymbol{\sigma}^2$. Given the equivalence between specifying priors on scalers for fixed and random effects noticed at the end of Section 3.2, the linear predictor (1) can be expressed as

$$\boldsymbol{\eta} = \mathbf{1}\beta_0 + \sum_{j=1}^{P+Q} \boldsymbol{\psi}_j,$$

where model components refer to fixed effects for $j \leq P$ and to random effects for $P < j \leq P + Q$. The sampling variance of the linear predictor, known as regression

variance in the context of linear regression models, is the random variable:

$$V_\eta = \sum_{j=1}^{P+Q} V_{\psi_j} + 2 \sum_{j=1}^{P+Q} \sum_{k<j} C_{\psi_j,\psi_k},$$

where terms C_{ψ_j,ψ_k} refer to the bilinear form $C_{\psi_j,\psi_k} = \boldsymbol{\psi_j}^\top \mathbf{M} \boldsymbol{\psi_k}/(n-1)$. Since variance components are a priori independent, $\mathbb{E}[C_{\psi_j,\psi_k}] = 0$, $\forall j \neq k$, and DSD priors guarantee that, $\mathbb{E}[V_\eta] = (P+Q)\mathbb{E}[\mathcal{V}]$. Concerning the posterior distributions of $V_{\psi_j}|\mathbf{y}$ and $C_{\psi_j,\psi_k}|\mathbf{y}$, in general, $\mathbb{E}[C_{\psi_j,\psi_k}|\mathbf{y}] \neq 0$ since the likelihood function induces dependence among model components: from an applied point of view, posterior analysis of variance/covariance decomposition can give insights about the explanatory power of each model component without ignoring their underlying relationships.

4 Simulation Exercise

The simulation study presents a comparative analysis between DSD priors and standard priors in terms of posterior sensitivity to the model's design and structure. Furthermore, we examine the influence of π_0 and q on posterior inference. We consider the data generating process

$$y_i \sim \mathcal{N}\big(g(x_i) = 5 + \sin(\pi x_i), \sigma_y^2\big), \quad i = 1, \ldots, 50;$$

where $\mathbf{x}$ is an equally spaced set of values in the interval $[-1; 1]$. A semi-parametric regression model implementing Bayesian P-splines (Lang and Brezger, 2004) is specified:

$$\boldsymbol{\eta} = \mathbf{1}\mu + \mathbf{Z}\boldsymbol{\gamma},$$

where the design matrix $\mathbf{Z}$ is a basis matrix of cubic B-splines over m equally spaced knots. For the spline coefficients $\boldsymbol{\gamma}$, a second-order RW prior is imposed by specifying the joint prior density

$$f\big(\boldsymbol{\gamma}|\sigma_\gamma^2\big) \propto \big(\sigma_\gamma^2\big)^{-(m-2)} \exp\left(-\frac{1}{2\sigma_{\gamma_j}^2} \sum_{i=1}^{m-2} (\gamma_i - 2\gamma_{i+1} + \gamma_{i+2})^2\right),$$

which leads to a precision matrix $\mathbf{K}_\gamma$ with rank $m-2$, requiring the inclusion of a linear trend. The performances of point and interval estimators of $g(x_i)$ are monitored by computing posterior means and 95% credible intervals.

We consider three different values for σ_y^2, to control the signal-to-noise ratio $\rho = \frac{\mathrm{Var}[\boldsymbol{\eta}]}{\mathrm{Var}[\boldsymbol{\eta}]+\sigma_y^2}$, which is set equal to $\rho \in \{0.25, 0.5, 0.75\}$. Moreover, we set a grid of values for m, ranging from 5 to 30 spaced by 5. The same LGM is fitted on $B = 500$ generated datasets, with different hyperpriors for the scale parameter σ_γ^2, distinguishing between standard priors (Half-Cauchy, the Half-t on the standard deviation, Inv-Gamma(10^{-3}, 10^{-3}), labeled with IG-J, and Inv-Gamma(1, 5×10^{-5}), labeled with IG-INLA, on the variance) and DSD priors for which hyperparameter values $q \in \{0.5, 1.5, 15\}$ and $\pi_0 \in \{0.1, 0.25, 0.5, 0.75\}$ are explored.

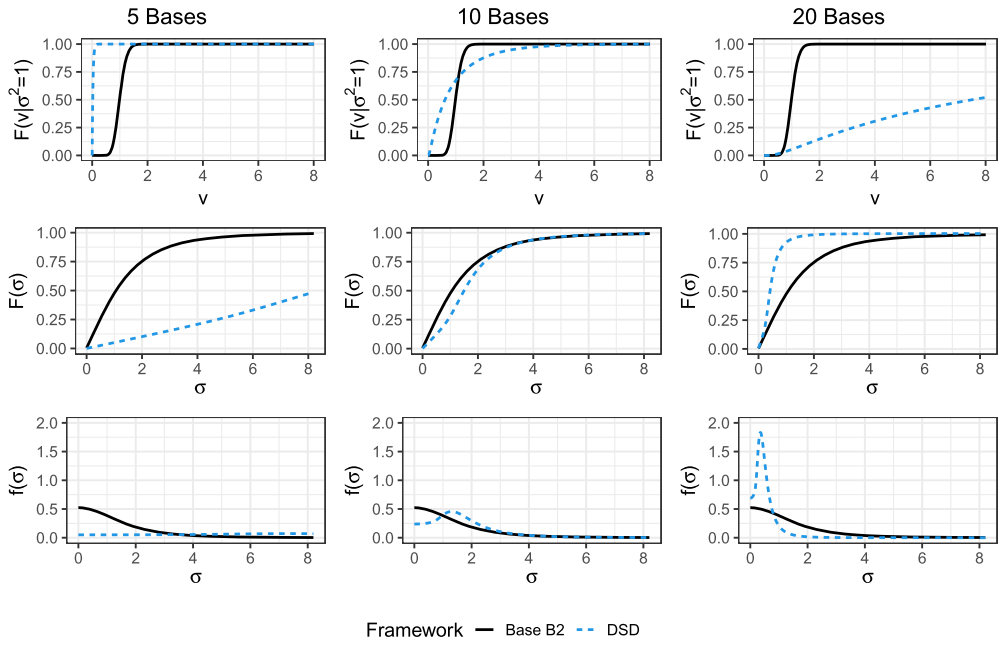

Figure 1: From top to bottom: CDFs of $V_\psi|\sigma^2 = 1$ (dashed line) and $\mathcal{V}|\sigma^2 = 1$ (solid line). CDFs and PDFs of implied priors on σ with a B2 prior or a DSD prior for σ^2.

Since a Gaussian model is assumed for the data, a prior for the residual variance must be set. In line with the indications provided within the `rstanarm` package (Goodrich et al., 2022), we specified an exponential distribution on the residual standard deviation, with rate parameter $1/s_y^2$.

Once the prior elicitation step described in Section 3.4 is carried out, and the hyperparameters b, p, and q are set, the goal is to impose a ${}_2\mathcal{F}_0(\alpha, \beta/b, p, q)$ prior distribution on V_ψ. We recall that it coincides with the marginal prior of V_ψ when an i.i.d effect is considered, with a $B2(b, p, q)$ hyperprior on the scale. However, different model specifications imply different effects structures, that deviate from the i.i.d. structure. Hence, the DSD prior modifies the base B2 distribution to preserve the ${}_2\mathcal{F}_0(\alpha, \beta/b, p, q)$ marginal prior on V_ψ.

Figure 1 shows how the DSD prior changes with respect to the selected number of bases m: the base B2 distribution is plotted up to quantile 0.999 determining the range of the x axis. We report the implied prior distributions on the standard deviation obtained by setting $\pi_0 = 0.5$ and $q = 1.5$; the priors have a positive finite density at 0 since $p = 0.5$. As it can be noticed from the first row of Figure 1, when $m = 5$ the CDF of $V_\psi|\sigma^2$ (dashed blue line) is steeper than the CDF of $\mathcal{V}|\sigma^2$ (black line). Consequently, a prior distribution that is flatter than the B2 density (see third row) is required on σ to obtain for V_ψ a marginal prior that matches the prior on $\mathcal{V}$. Increasing m reduces the steepness of the CDF of $V_\psi|\sigma^2$: when $m = 20$ the DSD prior induced on σ appears

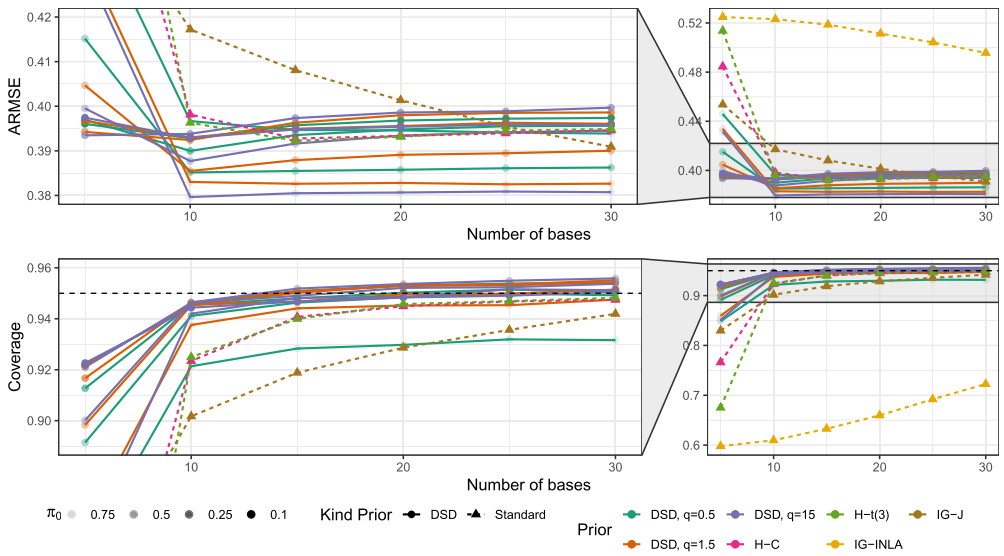

Figure 2: ARMSE and frequentist coverage with respect to the number of bases m for the estimates under the considered priors and $\rho = 0.25$.

to be peaked nearby 0, with the CDF that reaches the asymptote in 1 faster than the base $B2$ prior.

The performances of posterior means $\hat{\eta}_i = \mathbb{E}[\eta_i|\mathbf{y}]$ and 95% credible intervals are studied. As summary measures, the root mean squared error averaged over the curve (ARMSE) and the frequentist coverage are computed. They are defined as

$$\text{AE}^{(b)} = \sqrt{\frac{1}{50}\sum_{i=1}^{50}\left(\hat{\eta}_i^{(b)} - g(x_i)\right)^2}, \quad \text{ARMSE} = \frac{1}{B}\sum_{b=1}^{B}\text{AE}^{(b)},$$

$$\text{Coverage} = \frac{1}{B\cdot 50}\sum_{b=1}^{B}\sum_{i=1}^{50}\mathbf{1}\left(g(x_i) \in \left[\hat{\eta}_{i,0.025}^{(b)}, \hat{\eta}_{i,0.975}^{(b)}\right]\right),$$

where $\mathbf{1}(A)$ is an indicator function assuming value 1 if event A occurs and 0 otherwise. In addition, the superscript (b) indicates that the quantities refer to the generic iteration b, and $\hat{\eta}_{i,\mathcal{Q}}^{(b)}$ is the $\mathcal{Q}$-th posterior quantile of $\eta_i|\mathbf{y}$.

In Figure 2, the ARMSE and the frequentist coverage are reported. The main differences among the priors involved in the study arise when $m = 5$. In this case, the calibration carried out by DSD priors produces estimates characterized by a lower ARMSE and credible intervals with frequentist coverages closer to the nominal level. When m increases, models performances tend to converge quickly, with the exception of those obtained with the IG-INLA prior. Figure 2 focuses on the case $\rho = 0.25$, results for the remaining cases are reported in Figure S1 in the online Supplementary Material

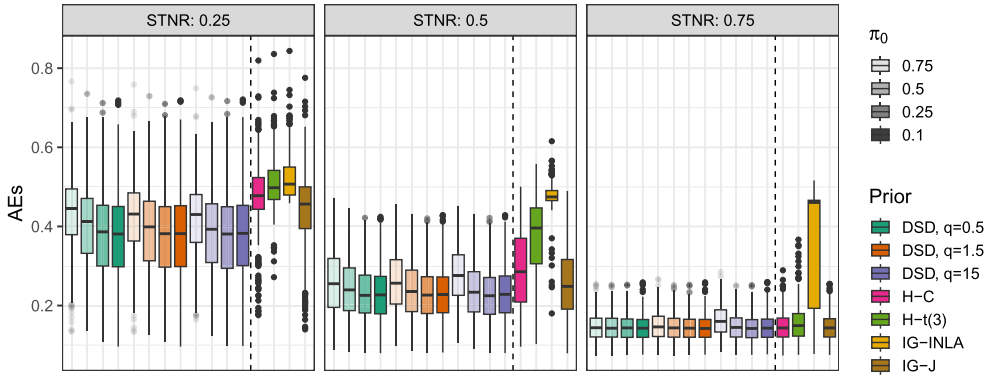

Figure 3: Monte Carlo distributions of AEs when $m = 5$. From left to right ρ is equal to $0.25, 0.5, 0.75$. The vertical dashed line separates results concerning DSD priors (on the left) from those under standard priors (on the right).

(Gardini et al., 2024a), where similar patterns can be detected.

To provide an idea of the uncertainty of the sampling distribution, the distributions of the average errors (AEs) of the curves are depicted in Figure 3 through boxplots. When $\rho \in \{0.25, 0.5\}$, DSD priors deliver lower AEs than standard priors. Furthermore, with the exception of $\pi_0 = 0.75$ for which higher errors are observed, the AEs distributions obtained with other values of π_0 are fairly stable. Lastly, results obtained for $\rho = 0.75$ show little variability, with the exception of the IG-INLA prior that produces markedly higher AEs values.

Lastly, the three panels of Figure 3 highlight that, when using DSD priors, changes in posterior inference are mainly due to π_0, while results are less reactive to changes of the q values.

5 Applications

In this section, we show two applications based on datasets available in the `R-INLA` package and analyzed in Rue and Held (2005) among others: the Munich rental guide and the Tokyo rainfall datasets.

5.1 Munich Rental Data

The dataset comprises $n = 2035$ observations from the 2003 Munich rental guide. The response variable is the rent price (per square meter in Euros); covariates comprise spatial location, floor space, year of construction, and a set of dichotomous variables describing house features such as presence/absence of central heating, bathroom, etc. To study prior sensitivity, we also consider a reduced dataset obtained by randomly sampling $n = 300$ observations. Following previous analyses, a Gaussian likelihood is

	m_j	$\widetilde{\alpha}$	$\widetilde{\beta}$	$\mathbb{E}[V_{\psi_j}\|\sigma^2_{\psi_j}=1]$	$\mathbb{V}[V_{\psi_j}\|\sigma^2_{\psi_j}=1]$
size	134	0.735	1.4×10^{-4}	5001.1	3.4×10^{7}
year	84	0.602	2.7×10^{-4}	2190.9	7.9×10^{6}
location	380	9.375	11.581	0.809	0.0698

Table 2: Characteristics of the model components sampling variances.

adopted and the linear predictor is specified as

$$\boldsymbol{\eta} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\psi}_{\text{year}} + \boldsymbol{\psi}_{\text{size}} + \boldsymbol{\psi}_{\text{loc}},$$

where

$$\boldsymbol{\psi_j} = \mathbf{Z}_j\boldsymbol{\gamma}_j; \quad \boldsymbol{\gamma}_j|\sigma^2_{\gamma_j} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2_{\gamma_j}\mathbf{K}_{\gamma_j}\big), \quad j \in \{\text{year, size, location}\}.$$

Both $\mathbf{K}_{\gamma_{\text{year}}}$ and $\mathbf{K}_{\gamma_{\text{size}}}$ are precision matrices of a continuous time second-order RW, hence their rank-deficiency is $\kappa = 2$. Since their null spaces are spanned by a first-order polynomial, corresponding linear terms are added to the linear predictor. This prior specification parallels the one proposed in Rue and Held (2005), being a sensible choice for estimating smooth covariate effects that we use to illustrate our approach. $\mathbf{K}_{\gamma_{\text{loc}}}$ is built as $\mathbf{D} - \mathbf{W}$ where $\mathbf{W}$ is the adjacency matrix of Munich districts and $\mathbf{D}$ is diagonal with i-th diagonal entry corresponding to the number of neighbors to the i-th district: being the graph connected, $\kappa = 1$ and no terms need to be added to the linear predictor. All random effects are constrained as described in Section 3.3. Concerning computational time, 25000 MCMC iterations require 32 seconds when adopting the Inverse Gamma prior and 39 seconds when adopting the DSD prior on a PC equipped with processor Intel Core i7-9750H (2.60 GHz).

Table 2 shows parameters of the conditional distributions $V_{\psi_j}|\sigma^2_{\gamma_j}$, where the last two columns report expectation and variance for $\sigma^2_{\gamma_j} = 1$. The same prior on scale parameters would result in sensibly different marginal distributions of V_{ψ_j}, with random effects built on RW structures being largely more variable a priori than the spatial random effect. The CDFs of DSD priors as a function of π_0 are shown in Figure S3 of the online Appendix: probability mass is distributed on far larger values for the spatial effect with respect to other components, in order to compensate for the aforementioned features of the conditional sampling variances.

In Figure 4, the posterior means of effects of size, year and location are reported for π_0 ranging from 0.05 to 0.95 and $q = 1.5$. In the complete dataset, prior sensitivity is limited, particularly in comparison with the reduced dataset, where year and location effects show marked sensitivity to π_0, due to reduced information. To highlight the changes due to π_0, sensitivity curves are reported in Figure 5, where relative $\mathbb{V}[\mathbb{E}[f(\mathbf{x}_j)|\mathbf{y}]]$ are shown for the three predictor effects $f(\mathbf{x}_j)$. All variances are divided by the component-wise smallest variance in order to emphasize relative changes. As expected, increasing π_0 encourages smoothing: location effect is the least sensitive to prior specification in the complete dataset, while the year effect shows some sensitivity in both cases.

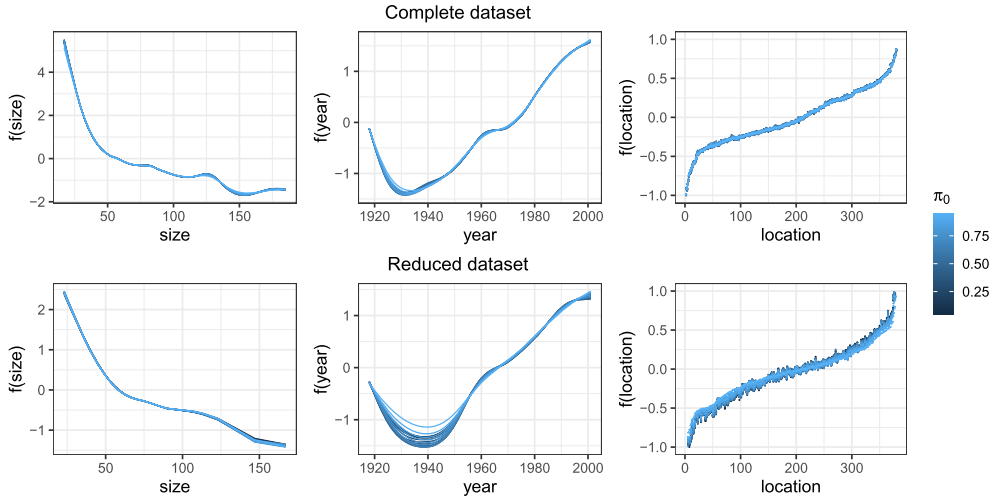

Figure 4: $\mathbb{E}[f(\mathbf{x}_j)|\mathbf{y}]$ of model components for complete and reduced datasets (q=1.5).

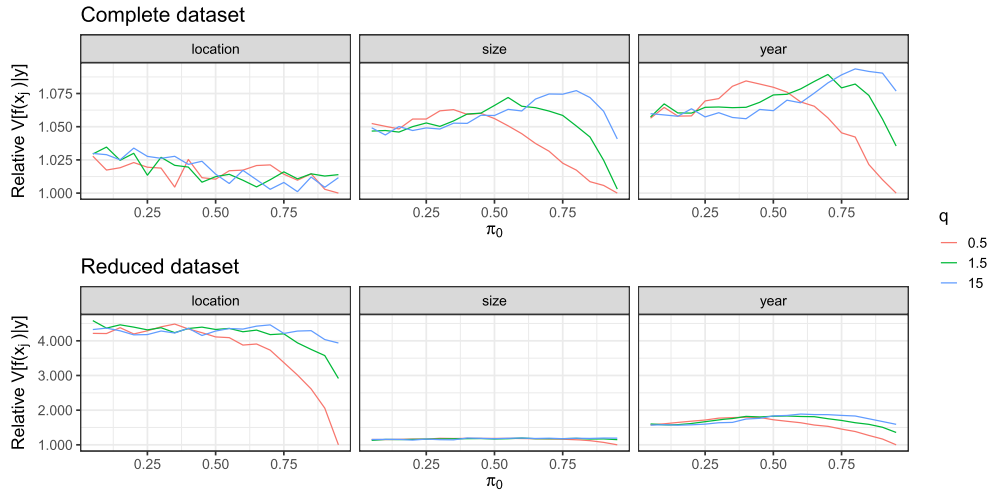

Figure 5: $\mathbb{V}[\mathbb{E}[f(\mathbf{x}_j)|\mathbf{y}]]$ divided by the component-wise maximum as a function of π_0.

Lastly, we study the decomposition of the linear predictor posterior variance $V_\eta|\mathbf{y}$, joining fixed effects. Draws from the posterior distribution of the variance explained by each model component $V_{\psi_j|\mathbf{y}}$ are obtained by summarising the MCMC chains. As an example, given a G-dimensional MCMC sample, posterior draws from $V_{\psi_{\text{size}}|\mathbf{y}}$ are obtained by computing the G variances of the sampled effects $\boldsymbol{\psi}^{(g)}_{\text{size}}$, i.e. $V^{(g)}_{\psi_{\text{size}}|y} = \text{var}(\boldsymbol{\psi}^{(g)}_{\text{size}})$, for $g = 1, \ldots, G$. Posterior covariances $C_{\psi_j,\psi_k} = \frac{\boldsymbol{\psi_j}^\top \mathbf{M} \boldsymbol{\psi_k}}{n-1}|\mathbf{y}$ are negligible,

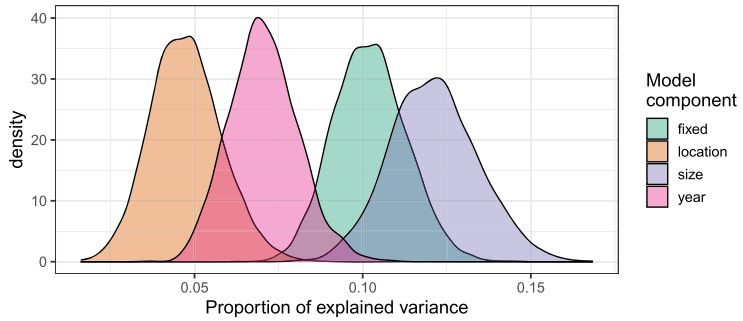

Figure 6: Proportion of explained variance by each model component captured by posterior distributions $V_{\psi_j}|\mathbf{y}$ divided by the data variance $s_y^2 = 6.08$. DSD priors with $q = 1.5$ and $\pi_0 = 0.9$.

hence posterior variances $V_{\psi_j}|\mathbf{y}$, whose distributions are shown in Figure 6, give clear-cut information on the contribution of each model component to the explained variability which receives the same prior weight by means of DSD priors. The size variable turns out to be the most important covariate for explaining rental prices, followed by fixed effects, year of construction and location.

5.2 Tokyo Rainfall Data

The dataset comprises $n = 731$ daily dichotomous observations which equal 1 if more than 1 mm of rainfall was recorded in Tokyo during 1983 and 1984, and 0 otherwise. The aim of the application is to estimate the rainfall probability p_t on a calendar day $t = 1, \dots, 366$. The Binomial likelihood is

$$y_t|p_t \sim \text{Bin}(p_t, n_t), \quad t = 1, \dots, 366,$$

where $n_t = 2$ for $t \neq 60$ and $n_{60} = 1$. The linear predictor is $g(\mathbf{p}) = \mathbf{1}\beta_0 + \boldsymbol{\psi}_{\text{day}}$ where

$$\boldsymbol{\psi}_{\text{day}} = \mathbf{Z}\boldsymbol{\gamma}_{\text{day}}; \quad \boldsymbol{\gamma}_{\text{day}}|\sigma^2_{\gamma_{\text{day}}} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2_{\gamma_{\text{day}}}\mathbf{K}_{\gamma_{\text{day}}}\big).$$

In this case, the random effect design matrix is $\mathbf{Z} = \mathbf{I}_{366}$ and the structure $\mathbf{K}_\gamma$ corresponds to a circular RW of order two, to introduce conditional dependence between the first and last day of the year (Rue and Held, 2005). Note that the IGMRF order is $\kappa = 1$, so the model needs only one linear constraint and no term needs to be added to the linear predictor.

We estimate the model adopting both the logit and probit link. In the former case, we set c in (17) as the pseudo-variance $c = \bar{y}^{-1}(1-\bar{y})^{-1} = 5.16$ (see Table 1 in Piironen and Vehtari, 2017), while $c = \bar{y}(1-\bar{y})/\phi(\Phi^{-1}(\bar{y}))^2 = 1.82$ for the probit link. This leads to $b = 26.5$ and $b = 9.34$ for the logit and probit links respectively, when $q = 1.5$ and $p = 0.5$ as suggested in Section 3.4 and $\pi_0 = 0.5$. Concerning computational time, 25000 MCMC iterations with probit link require 29 seconds when adopting the Inverse

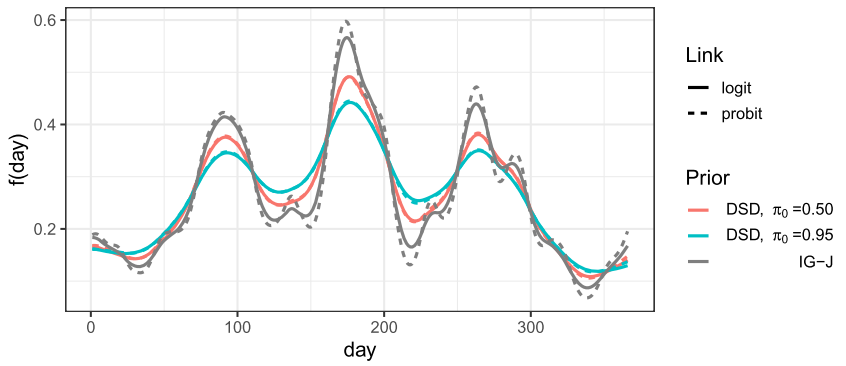

Figure 7: Fitted probabilities with respect to calendar days.

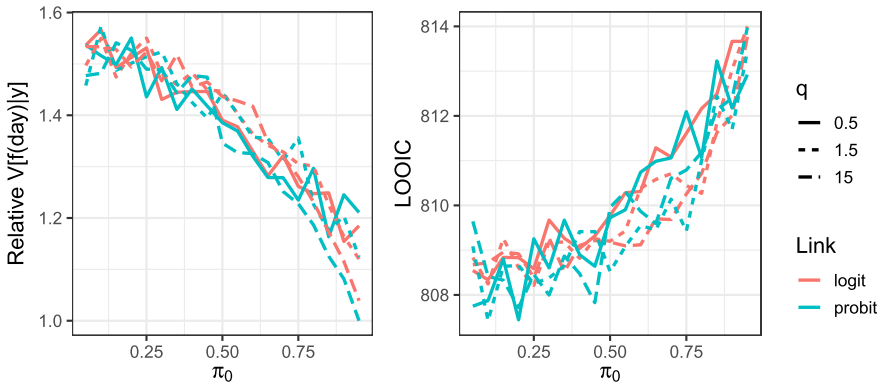

Figure 8: $\mathbb{V}[\mathbb{E}[f(\text{day})|\mathbf{y}]]$ divided by the maximum as a function of π_0 (left panel). Leave One Out Information Criterion as a function of π_0 (right panel).

Gamma prior and 34 seconds when adopting the DSD prior on a PC equipped with processor Intel Core i7-9750H (2.60 GHz). At the moment, the `C++` code for sampling the logit model is highly inefficient, being demanding in terms of computational time.

The merit of assigning different priors in the probit and the logit scale is highlighted in Figure 7, where one can see that adopting the same IG-J prior independently on the link delivers a sensibly different smoothness of the fitted probabilities as a function of calendar day. On the other hand, DSD priors show limited variation with respect to the type of link function, being the smoothness of the predicted probabilities mainly affected by π_0. In this application, posterior inference shows a marked sensitivity to π_0, as shown in the sensitivity curves reported in the left panel of Figure 8: this is not surprising because of the limited amount of information available to estimate a smooth function of time.

In the right panel of Figure 8 the leave-one-out cross-validation information criterion (LOOIC, Vehtari et al., 2017) is reported. LOOIC serves as a model selection criterion,

giving preference to models demonstrating lower values of the indicator. It can be seen that LOOIC increases as the amount of shrinkage (determined by π_0) increases. We observed a similar behavior in the Munich rental data application (see Figure S2), particularly with reference to the reduced dataset. This suggests that extreme values of π_0 can give rise to excessive smoothing resulting in worse model performances. On the basis of these considerations, we suggest to set $\pi_0 = 0.5$ as a default value, completing the default setting sketched in Section 3.4.

6 Concluding Remarks

In this paper, a unified strategy for prior specification of scale parameters in latent Gaussian models has been proposed. This is a debated problem in the literature about Bayesian Hierarchical models and several interesting proposals have been developed to date. In our opinion, the lack of agreement about sensible default strategies for prior specification in this kind of models is due to three aspects that hinder clear-cut interpretation of the priors on parameters that act as mere scalers: design matrices, structure matrices, and linear constraints. From an applied point of view, we believe that the most important phase of model specification concerns indeed these three components which contain: auxiliary information on the phenomenon under study (design), prior beliefs on the correlation structure/smoothness of model components (structure) and precautions needed for model identifiability or, possibly, for customizing parameters interpretability (linear constraints). Scale parameters are far less relevant in model specification, but they require careful prior specification in order to obtain a sensible allocation of prior variability to model components. As shown in Section 3, this can be done by taking into account the whole model architecture, comprising the link function. From a methodological point of view, this approach is in line with the discussion provided in Gelman et al. (2017), particularly when the authors discuss the role of the prior in *generative and predictive* modeling. Indeed, the π_0 parameter in DSD priors can be interpreted as a measure of plausibility of observed data under the model: for this reason, we find it sensible to set π_0 on the basis of the data variability when a Gaussian likelihood is concerned or on pseudo-variance in generalized linear models.

DSD priors operate by transforming the base B2(b, p, q) prior to accommodate all model features impacting the marginal variance of the linear predictor. By default, we set parameter $p = 1/2$ for two reasons. Firstly, this selection ensures a positive finite density at zero on the standard deviation of the model component, thereby assigning positive probability to models that exclude the component, a feature also suggested in the development of the theory regarding Penalized Complexity Priors (Simpson et al., 2017). Secondly, setting $p = 1/2$ guarantees the existence of the DSD prior, as stated in Theorem 3.1. Parameter q governs the right tail of the prior distribution; we recommend setting $q = 1.5$, corresponding to a Half-t distribution with 3 degrees of freedom on the standard deviation scale. Regardless of the observed low sensitivity to this parameter in our simulations and applications, we consider its impact on posterior inference to be less critical compared to the scale parameter b. This last parameter is specified as a function of π_0, which establishes a crucial link between the prior distribution and the data being modeled, as the data variance corresponds to the π_0-th quantile

of the model component's marginal variance. As a default value, we set $\pi_0 = 0.5$ in the `DSDprior` package, meaning that data variance corresponds to the median of the prior marginal sampling variance of each model component. Regardless that it has been shown to be a sensible value in our applications, we recognize that this choice lacks a particularly compelling rationale; informally, it suggests that observed data are 'reasonably' generated by the data generating process implied by the prior. In our opinion, it is not possible to prescribe default values that universally suit all diverse applications and, of course, default values are not able to reflect different modeler's prior beliefs. The true merit of π_0 lies in its consistency of interpretability across various models, facilitating the expression of prior beliefs and serving as a natural choice for investigating the sensitivity of posterior inference to prior choices.

Concerning usability and easiness of application of latent Gaussian models, we think that DSD priors can be a useful tool for practitioners to be used in disparate situations comprising spatial, temporal, spatio-temporal, ANOVA, semi-parametric models and in general all models that can be cast coherently with the models summarised in Table 1. One limitation of the proposed approach lies in its applicability only to models characterized by fixed structure matrices. Extension to models containing parameterized structure matrices poses several challenges related to integration over parameters that, being contained within the structure matrix, greatly complicate the study of the distribution of quadratic forms involved in the analysis. Nonetheless, extensions to more general models will be the subject of future research.

Besides the practical aspects, another goal of the paper is to deepen the theory behind latent Gaussian models, trying to characterize conditional and marginal sampling variances of linear predictor components. These quantities provide a meaningful interpretation of scale parameters, useful in the prior elicitation step. These developments allow contextualizing previous contributions in the field like those by Sørbye and Rue (2014) and Klein and Kneib (2016). Indeed, the scaling procedure by the former can be seen as an attempt of removing the impact of the structure in the conditional expectation of the sampling variance, as deducible by results in Section 3.1. On the other hand, finding a prior that solves the integral equation in (12) is in line with the proposal of scale dependent priors by Klein and Kneib (2016), where the prior on scale parameters is retrieved numerically, relying on a different synthesis of the effect dispersion and incorporating the design through arbitrary covariate patterns. In this sense, taking the sampling variance of the effect as focal quantity leads us to meaningful distributional results by resorting to the theory of quadratic forms.

As regards the allocation of prior variability to model components, all the developments discussed in the paper implicitly assume that the modeler has no preference for some specific components, all being assigned the same prior variability. A potentially useful extension that we plan to develop in future work is to allow unbalancing of prior variability assigned to model components by introducing some a priori weighting system: to this aim, we find it appealing to resort to the approach proposed in Fuglstad et al. (2020).

Funding

The work of Aldo Gardini was partially supported by MUR on funds FSE REACT EU - PON R&I 2014-2020 and PNR (D.M. 737/2021) for the RTDA_GREEN project (title: "Modelli statistici per lo studio della convergenza spaziale verso la transizione verde", J41B21012140007).

Supplementary Material

Supplementary material to: Design and Structure Dependent Priors for Scale Parameters in Latent Gaussian Models (DOI: 10.1214/24-BA1454SUPPA; .pdf). A file containing an introduction to quadratic forms, the technical details of the proofs and additional outputs from the simulation study and the applications.

R code (DOI: 10.1214/24-BA1454SUPPB; .zip). A zipped folder containing the binary file of the `DSDprior` R package with the functions useful to implement an MCMC sampler to fit models with DSD priors. Additional scripts containing the code to run the examples in the applications are provided.

References

Aguilar, J. E. and Bürkner, P.-C. (2023). "Intuitive joint priors for Bayesian linear multilevel models: The R2D2M2 prior." *Electronic Journal of Statistics*, 17(1): 1711–1767. MR4609453. doi: https://doi.org/10.1214/23-ejs2136. 6

Bernardinelli, L., Clayton, D., and Montomoli, C. (1995). "Bayesian estimates of disease maps: how important are priors?" *Statistics in Medicine*, 14(21-22): 2411–2431. 5

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43(1): 1–20. MR1105822. doi: https://doi.org/10.1007/BF00116466. 5

Box, G. E. P. (1954). "Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification." *The Annals of Mathematical Statistics*, 25(2): 290–302. MR0061787. doi: https://doi.org/10.1214/aoms/1177728786. 6, 9, 10

Fuglstad, G. A., Hem, I. G., Knight, A., Rue, H., and Riebler, A. (2020). "Intuitive joint priors for variance parameters." *Bayesian Analysis*, 15(4): 1109–1137. MR4171145. doi: https://doi.org/10.1214/19-BA1185. 6, 7, 23

Gardini, A., Greco, F., and Trivisano, C. (2022). "The Mellin Transform to manage quadratic forms in normal random variables." *Journal of Computational and Graphical Statistics*, 31(4): 1416–1425. MR4513398. doi: https://doi.org/10.1080/10618600.2022.2034639. 3, 10

Gardini, A., Greco, F., and Trivisano, C. (2024a). "Supplementary Material for "Design and structure dependent priors for scale parameters in latent Gaussian models"."

Bayesian Analysis. doi: https://doi.org/10.1214/24-BA1454SUPPA. 6, 8, 9, 10, 11, 17

Gardini, A., Greco, F., and Trivisano, C. (2024b). "Supplementary Material for "Design and structure dependent priors for scale parameters in latent Gaussian models"." *Bayesian Analysis*. doi: https://doi.org/10.1214/24-BA1454SUPPB. 11

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). "Efficient parametrisations for normal linear mixed models." *Biometrika*, 82(3): 479–488. MR1366275. doi: https://doi.org/10.1093/biomet/82.3.479. 2

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper)." *Bayesian Analysis*, 1(3): 515–534. MR2221284. doi: https://doi.org/10.1214/06-BA117A. 4

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis (3rd ed.)*. Chapman and Hall/CRC. MR3235677. 3, 13

Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press. MR2262986. doi: https://doi.org/10.1137/040607964. 5

Gelman, A., Simpson, D., and Betancourt, M. (2017). "The prior can often only be understood in the context of the likelihood." *Entropy*, 19(10). 2, 22

Goicoa, T., Adin, A., Ugarte, M., and Hodges, J. (2018). "In spatio-temporal disease mapping models, identifiability constraints affect PQL and INLA results." *Stochastic Environmental Research and Risk Assessment*, 32(3): 749–770. MR3732861. doi: https://doi.org/10.1016/j.spasta.2017.04.002. 11

Goodrich, B., Gabry, J., Ali, I., and Brilleman, S. (2022). "rstanarm: Bayesian applied regression modeling via Stan." R package version 2.21.3. URL https://mc-stan.org/rstanarm/ 13, 15

Klein, N. and Kneib, T. (2016). "Scale-dependent priors for variance parameters in structured additive distributional regression." *Bayesian Analysis*, 11(4): 1071–1106. MR3545474. doi: https://doi.org/10.1214/15-BA983. 2, 5, 11, 12, 23

Knorr-Held, L. and Rue, H. (2002). "On block updating in Markov random field models for disease mapping." *Scandinavian Journal of Statistics*, 29(4): 597–614. MR1988414. doi: https://doi.org/10.1111/1467-9469.00308. 11

Lang, S. and Brezger, A. (2004). "Bayesian P-splines." *Journal of Computational and Graphical Statistics*, 13(1): 183–212. MR2044877. doi: https://doi.org/10.1198/1061860043010. 14

Olver, F., Lozier, D., Boisvert, R., and Clark, C. (2010). *The NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY. MR2723248. 8, 9

Perez, M. E., Pericchi, L. R., and Ramirez, I. C. (2017). "The scaled beta2 distribution as a robust prior for scales." *Bayesian Analysis*, 12(3): 615–637. MR3655869. doi: https://doi.org/10.1214/16-BA1015. 4, 12

Piironen, J. and Vehtari, A. (2017). "Sparsity information and regularization in the horseshoe and other shrinkage priors." *Electronic Journal of Statistics*, 11(2): 5018–5051. MR3738204. doi: https://doi.org/10.1214/17-EJS1337SI. 13, 20

Polson, N. G. and Scott, J. G. (2012). "On the half-Cauchy prior for a global scale parameter." *Bayesian Analysis*, 7(4): 887–902. MR3000018. doi: https://doi.org/10.1214/12-BA730. 4

Riebler, A., Sørbye, S. H., Simpson, D., Rue, H., Lawson, A. B., Lee, D., and MacNab, Y. (2016). "An intuitive Bayesian spatial model for disease mapping that accounts for scaling." *Statistical Methods in Medical Research*, 25(4): 1145–1165. MR3541089. doi: https://doi.org/10.1177/0962280216660421. 5

Ruben, H. (1962). "Probability content of regions under spherical normal distributions, IV: The distribution of homogeneous and non-homogeneous quadratic functions of normal variables." *The Annals of Mathematical Statistics*, 33(2): 542–570. MR0137189. doi: https://doi.org/10.1214/aoms/1177704580. 6

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall. MR2130347. doi: https://doi.org/10.1201/9780203492024. 5, 11, 17, 18, 20

Rue, H., Martino, S., and Chopin, N. (2009). "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2): 319–392. MR2649602. doi: https://doi.org/10.1111/j.1467-9868.2008.00700.x. 5

Schrödle, B. and Held, L. (2011). "Spatio-temporal disease mapping using INLA." *Environmetrics*, 22(6): 725–734. MR2843139. doi: https://doi.org/10.1002/env.1065. 11

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). "Penalising model component complexity: A principled, practical approach to constructing priors." *Statistical Science*, 32(1): 1–28. MR3634300. doi: https://doi.org/10.1214/16-STS576. 4, 12, 22

Sørbye, S. H. and Rue, H. (2014). "Scaling intrinsic Gaussian Markov random field priors in spatial modelling." *Spatial Statistics*, 8(C): 39–51. MR3326820. doi: https://doi.org/10.1016/j.spasta.2013.06.004. 5, 7, 23

Vehtari, A., Gelman, A., and Gabry, J. (2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and computing*, 27(5): 1413–1432. MR3647105. doi: https://doi.org/10.1007/s11222-016-9696-4. 21

Wakefield, J. (2007). "Disease mapping and spatial regression with count data." *Biostatistics*, 8(2): 158–183. 2, 12

Yanchenko, E., Bondell, H. D., and Reich, B. J. (2021). "The R2D2 prior for generalized linear mixed models." *arXiv preprint arXiv:2111.10718*. 6

Yanchenko, E., Bondell, H. D., and Reich, B. J. (2023). "Spatial regression modeling via the R2D2 framework." *Environmetrics*, e2829. 6

Zhang, Y. D., Naughton, B. P., Bondell, H. D., and Reich, B. J. (2020). "Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior." *Journal of the American Statistical Association*, 117(58): 862–874. MR4436318. doi: https://doi.org/10.1080/01621459.2020.1825449. 6