Alma Mater Studiorum
Università di Bologna

DOTTORATO DI RICERCA IN
SCIENZE STATISTICHE

CICLO XXXII

Settore Concorsuale:	13/D1
Settore Scientifico Disciplinare:	SECS-S/01

Bayesian inference for quantiles and conditional means in log-normal models

Presentata da: Aldo Gardini

Coordinatrice Dottorato:
Prof.ssa Alessandra Luati

Supervisor:
Prof. Carlo Trivisano

Esame Finale Anno 2020

Abstract

The main topic of the thesis is the proper execution of a Bayesian inference if log-normality is assumed for data. In fact, it is known that a particular care is required in this context, since the most common prior distributions for the variance in log scale produce posteriors for the log-normal mean which do not have finite moments. Hence, classical summary measures of the posterior such as expectation and variance cannot be computed for these distributions.

The thesis is aimed at proposing solutions to carry out Bayesian inference inside a mathematically coherent framework, focusing on the estimation of two quantities: log-normal quantiles (first part of the thesis) and conditioned expectations under a general log-normal linear mixed model (second part of the thesis). Moreover, in the latter section, a further investigation on a unit-level small area models is presented, considering the problem of estimating the well-known log-transformed Battese, Harter and Fuller model in the hierarchical Bayes context.

Once the existence conditions for the moments of the target functionals posterior are proved, new strategies to specify prior distributions are suggested. Then, the frequentist properties of the deduced Bayes estimators and credible intervals are evaluated through accurate simulations studies: it resulted that the proposed methodologies improve the Bayesian estimates under naive prior settings and are satisfactorily competitive with the frequentist solutions available in the literature. To conclude, applications of the developed inferential strategies are illustrated on real datasets.

The work is completed by the implementation of an `R` package named `BayesLN` which allows the users to easily carry out Bayesian inference for log-normal data.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The use of the logarithmic transformation in statistics has a long tradition. One of its main applications is the normalization of samples for which the Gaussian assumption is unreliable in the original scale. In fact, after the introduction of the analysis of variance method, whose starting point could be considered the paper by Fisher and Mackenzie (1923), the necessity to generalize this revolutionary technique to non-normal data emerged. In this sense, the log-transformation appears in the paper by Cochran (1938) among other transformations aimed at making experimental data suitable to apply the analysis of variance method, i.e. homoscedasticity and normality (with a particular focus on skewed distributions). In this paper, Cochran writes: "*The transformation to logs. equalizes the variance when it is proportional to the square of the mean; it is thus a much more powerful transformation than the square root or the inverse sine*".

Another example of early use of the logarithmic transformation by applied scientists is the paper by Williams (1937). He focuses on the fact that log-transforming the data is appealing when the the goal of the statistical analysis is the estimation of the geometric mean: back-transforming the arithmetic mean estimated on the log-transformed data, an estimate of the geometric mean of the original data is obtained (McAlister, 1879). Few years later, a different perspective of the idea of data transformation is provided in the paper by Finney (1941), that could be seen as the starting point of the inference on log-normal distribution. In effect, he noted that, if the log-transformation is performed, then a crucial step of the inferential procedure is represented by the back-transformation to the original data scale. In particular, he presented the first proposal of an efficient estimator for the arithmetic mean of the original data, exploiting the properties of the log-normal distribution.

The popularity of the log-transformation further increased thanks to the well known procedure introduced by Box and Cox (1964). In fact, the logarithm represents a particular case of the proposed data transformation algorithm.

Unfortunately, in applied sciences (but also among statisticians), there is a huge misunderstanding about the adoption of the logarithmic transformation in data analysis. Probably, the main source of confusion is represented by the following wrong procedure: *(1) transforming the data, (2) applying the well-known normal methods on the log-transformed sample and*

then (3) naively back-transform the results to the original data scale, ignoring basic concepts as Jensen's inequality. On the other hand, if the real interest of the analysis is to produce inference about key quantities of the original data scale (i.e. the scale of the untransformed data), then the idea of transformation should be abandoned, passing to the more coherent idea of carrying out a proper and careful inference on skewed data *whose logarithm is normally distributed* (Finney, 1941), and the log-normal (Crow and Shimizu, 1987) distribution might represent a convenient assumption.
This particular distribution is frequently used in several applied fields like economics, environmental sciences, biostatistics and engineering, in order to analyse different kind of data (Limpert et al., 2001).

1.1 The log-normal distribution

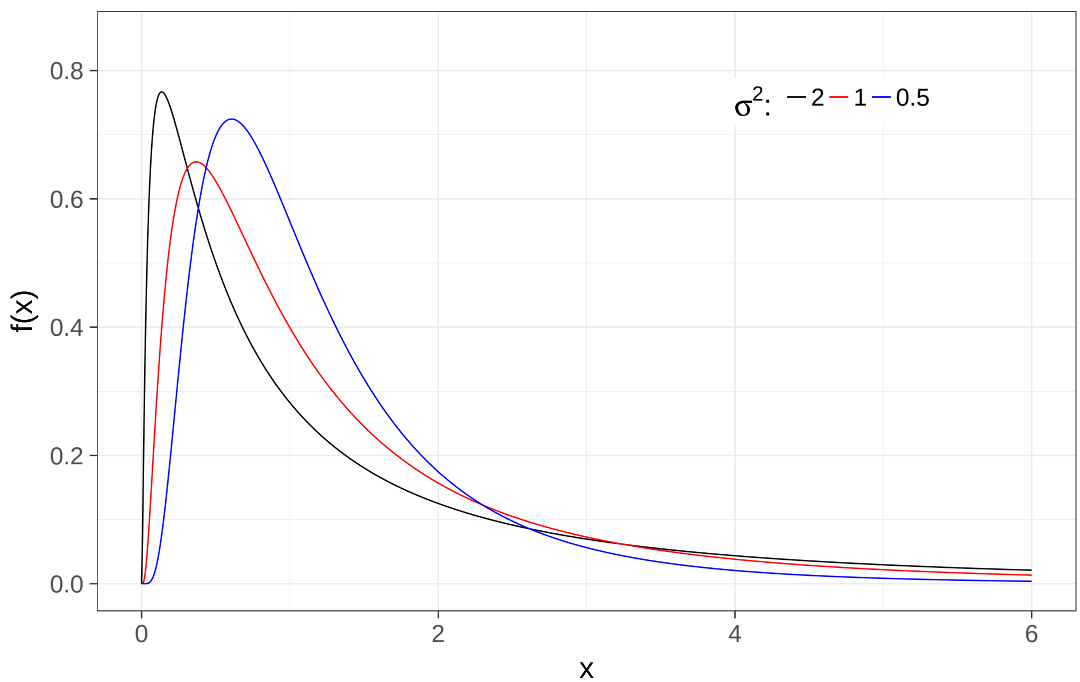

Figure 1.1: Plot of the density function of a log-normal distribution with $\xi = 0$ and different values of σ^2.

A brief overview of the log-normal distribution is provided in order to introduce the protagonist of this thesis and to fix the notation. If a random variable X is assumed normally distributed:

$$X \sim \mathcal{N}\left(\xi, \sigma^2\right), \tag{1.1}$$

then, its exponentiation $Z = \exp\{X\}$ is a positive random variable, which is log-normally distributed:

$$Z \sim \log \mathcal{N}(\xi, \sigma^2). \tag{1.2}$$

The probability density function (figure 1.1) is:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}\sigma z} \exp\left\{-\frac{1}{2\sigma^2}(\log z - \xi)^2\right\}, \qquad z > 0. \tag{1.3}$$

The family of functionals which includes the most important quantities that characterize the distribution is:

$$\theta_{a,b} = \exp\left\{a\xi + b\sigma^2\right\}. \tag{1.4}$$

It provides the arithmetic mean if $a = 1,\ b = 0.5$; the median when $a = 1,\ b = 0$ and the mode with $a = 1,\ b = -1$. More generally, each raw moment can be expressed choosing a couple of values for a and b.
On the other hand, the functional that defines the quantiles of the log-normal distribution is the following:

$$\theta_p = \exp\left\{\xi + \Phi^{-1}(p)\sigma\right\}, \tag{1.5}$$

where $\Phi^{-1}(p)$ is the inverse of the standard normal cumulative distribution function, i.e. the p-th quantile of the standardized Gaussian distribution which corresponds to probability p.
To complete the general characterization of the distribution, the variance of a log-normal random variable is:

$$\mathbb{V}[Z] = \left(e^{\sigma^2} - 1\right) e^{2\xi + \sigma^2}, \tag{1.6}$$

from which follows that the coefficient of variation does not depend on the mean in the log-scale ξ:

$$CV[Z] = \sqrt{e^{\sigma^2} - 1}. \tag{1.7}$$

Finally, another useful quantity to report is the distribution skewness:

$$\left(e^{\sigma^2} + 2\right)\sqrt{e^{\sigma^2} - 1}. \tag{1.8}$$

1.2 Log-normal distribution in Bayesian inference: a motivating example

As outlined before, one of the most critical points of making inference with the log-normal assumption is the back-transformation of the results which are obtained in the Gaussian framework to the original data scale, without neglecting the achievement of efficient estimators.
The study of appropriate and efficient estimators for crucial quantities related to the log-normal distribution is an active research field in statistics. A lot of papers about the estimation of the mean have been published (Zhou, 1998; Shen et al., 2006); also the quantiles estimation has received some attention (Longford, 2012). Moreover, in many applications linear mixed models with the response variable log-normally distributed are used, and the

problem of estimating quantities in the original data scale has not been extensively faced, yet.
Even if it could be supposed that these inferential issues might be easily overcame in the Bayesian framework sampling directly from the posterior distributions of the target functional, other problems related to the posteriors obtained with the widespread normal conjugate analysis are often ignored.
In effect, proposing the usual Bayes estimator under the quadratic loss function (i.e. the posterior mean), the finiteness of the posterior moments must be assured at least up to the second order, to obtain the posterior variance too. This step is often overlooked, but it is crucial to perform a coherent Bayesian analysis. Furthermore, this issue could be masked if the estimation is performed through MCMC methods (Sun and Speckman, 2005; Ghosh et al., 2018).
When an improper prior is fixed, a lot of care is usually taken in the properness of the posterior distribution. However, the Bayes estimators of log-normal functionals do not exist with the usual priors, both improper and proper (like the inverse gamma). For example, in the context of the log-normal mean, this issue was highlighted by Zellner (1971), and interesting solutions were proposed by Rukhin (1986) and Fabrizi and Trivisano (2012).
To provide the general idea of the way in which the non-existence of the posterior moments affects the usual inference based on MCMC methods, a simple simulation exercise is shown as motivating example. Firstly, a random sample from a log-normal distribution with parameters $\xi = 2$ and $\sigma^2 = 1$ is considered as dataset and samples from the posterior of the target quantities are generated using different tools: `JAGS` through `rjags` (Plummer, 2016), OpenBUGS (Spiegelhalter et al., 2007), `Stan` (Carpenter et al., 2017) and implementing the Gibbs samples exploiting the conjugacy of the estimated model, since the conjugate normal-inverse gamma prior is assumed at this stage.
For each method, two independent posterior samples of size $1,000,000$ are generated, after a burn-in period of $100,000$ iterations. The quantities included in the exercise are the sample mean, the median and the quantiles that correspond to $p = 0.1$ and $p = 0.9$. All the traceplots of the chains evidenced the convergence to a unique stationary distribution. The mean and standard deviation of the posterior distributions are reported in table 1.1 for three different sample sizes n.
Then, the simple one way ANOVA random effect model with J balanced groups having sample size n_g is considered for the response variable logarithm:

$$\begin{aligned} &\log(y_{ij}) = \mu + \nu_j + \varepsilon_{ij}, \quad i = 1, ..., n_g, \quad j = 1, ..., J; \\ &\nu_j \sim \mathcal{N}(0, \tau^2); \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2). \end{aligned} \tag{1.9}$$

In this case, to generate the toy dataset, the parameters are fixed as $\mu = 1$, $\sigma^2 = 1$ and $\tau^2 = 1$. Following the indications in Gelman (2006), half-t priors with 3 degrees of freedom are specified for the scale parameters:

$$\begin{aligned} &\sigma \sim \text{half} - t_3, \\ &\tau \sim \text{half} - t_3. \end{aligned} \tag{1.10}$$

n		JAGS				OpenBUGS			
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
10	$\theta_{1,0.5}$	23.9	3×10^3	19.3	1×10^3	23.3	5×10^3	2×10^3	2×10^7
	$\theta_{0.5}$	11.5	3.3	11.5	3.2	11.5	3.3	11.5	3.2
	$\theta_{0.1}$	4.1	1.4	4.1	1.4	4.1	1.4	4.1	1.4
	$\theta_{0.9}$	34.9	86.4	34.8	36.0	34.9	85.5	34.9	86.9
15	$\theta_{1,0.5}$	14.2	16.5	14.2	4.0	14.2	3.9	14.2	5.6
	$\theta_{0.5}$	10.5	2.1	10.5	2.1	10.5	2.1	10.5	2.1
	$\theta_{0.1}$	4.1	1.1	4.1	1.1	4.1	1.1	4.1	1.1
	$\theta_{0.9}$	27.8	9.3	27.8	9.3	27.8	9.2	27.9	9.4
20	$\theta_{1,0.5}$	12.7	2.7	12.7	2.8	12.7	2.8	12.7	2.8
	$\theta_{0.5}$	9.5	1.6	9.5	1.6	9.5	1.6	9.5	1.6
	$\theta_{0.1}$	3.7	0.8	3.7	0.8	3.7	0.8	3.7	0.8
	$\theta_{0.9}$	25.0	6.6	25.0	6.6	25.0	6.6	25.0	6.6
		Stan				Gibbs sampler			
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
10	$\theta_{1,0.5}$	2×10^3	2×10^6	17.5	200.3	19.3	925.1	33.8	2×10^4
	$\theta_{0.5}$	11.5	3.2	11.5	3.2	11.5	3.2	11.5	3.2
	$\theta_{0.1}$	4.1	1.4	4.1	1.4	4.1	1.4	4.1	1.4
	$\theta_{0.9}$	35.0	79.6	34.7	25.0	34.8	36.0	34.8	30.2
15	$\theta_{1,0.5}$	14.2	4.0	14.2	4.2	14.2	4.0	14.2	4.1
	$\theta_{0.5}$	10.5	2.1	10.5	2.1	10.5	2.1	10.5	2.1
	$\theta_{0.1}$	4.1	1.1	4.1	1.1	4.1	1.1	4.1	1.1
	$\theta_{0.9}$	27.9	9.3	27.8	9.4	27.8	9.3	27.8	9.2
20	$\theta_{1,0.5}$	12.7	2.8	12.7	2.8	12.7	2.8	12.7	2.7
	$\theta_{0.5}$	9.5	1.6	9.5	1.6	9.5	1.6	9.5	1.6
	$\theta_{0.1}$	3.7	0.8	3.7	0.8	3.7	0.8	3.7	0.8
	$\theta_{0.9}$	25.0	6.6	25.0	6.6	25.0	6.6	25.0	6.6

Table 1.1: Results of the MCMC exercise involving the estimation of mean ($\theta_{1,0.5}$), median ($\theta_{0.5}$) and quantiles ($\theta_{0.1}$, $\theta_{0.9}$) for the toy example.

In this framework, if the goal of the analysis is to predict the group-level mean for the original data scale, the functional considered is:

$$\theta_c(\nu_l) = \exp\left\{\mu + \nu_j + \frac{\sigma^2}{2}\right\}, \quad j = 1, ..., J. \tag{1.11}$$

On the other hand, if the global mean is the key quantity of the analysis, then the functional to estimate is:

$$\theta_m = \exp\left\{\mu + \frac{\sigma^2 + \tau^2}{2}\right\}. \tag{1.12}$$

Also in this case, two independent posterior samples are drawn for the target functionals and the posterior means and standard deviations are reported in table 1.2.

		JAGS				Stan			
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
$n_g = 3$ $J = 4$	$\theta_c(\nu_1)$	8.3	900.1	10.5	2×10^3	9.3	817.9	785.3	7×10^5
	$\theta_c(\nu_2)$	4.3	1×10^3	14.5	9×10^3	3.6	308.2	3×10^4	3×10^7
	$\theta_c(\nu_3)$	36.6	329.5	50.0	8×10^3	39.4	2×10^3	5×10^4	5×10^7
	$\theta_c(\nu_4)$	9.8	357.1	17.2	6×10^3	11.5	983.3	2×10^5	2×10^8
	θ_m	2×10^{14}	1×10^{17}	1×10^{23}	9×10^{25}	3×10^{129}	2×10^{132}	1×10^{72}	1×10^{75}
$n_g = 5$ $J = 4$	$\theta_c(\nu_1)$	3.8	37.4	3.8	3.4	3.7	3.0	3.8	3.6
	$\theta_c(\nu_2)$	6.0	20.9	6.0	4.3	6.0	5.1	6.0	4.3
	$\theta_c(\nu_3)$	21.8	106.9	21.7	32.3	21.6	14.3	21.7	14.8
	$\theta_c(\nu_4)$	3.4	9.3	3.4	3.2	3.4	2.7	3.3	2.8
	θ_m	3×10^{28}	3×10^{31}	5×10^8	4×10^{11}	4×10^{67}	4×10^{70}	7×10^{125}	Inf
$n_g = 3$ $J = 10$	$\theta_c(\nu_1)$	5.2	2.7	5.2	2.7	5.2	2.7	5.2	2.7
	$\theta_c(\nu_5)$	3.2	1.9	3.2	1.9	3.2	1.8	3.2	1.8
	$\theta_c(\nu_10)$	3.2	1.8	3.2	1.8	3.2	1.8	3.2	1.8
	θ_m	7.7	21.3	7.7	9.2	8.1	250.3	7.8	43.9

Table 1.2: Results of the MCMC exercise involving the estimation of global mean (θ_m) and group means ($\theta_c(\nu_l)$) under a one way random effect ANOVA model, considering the toy dataset.

By looking at the tables containing the simulation exercise results, the motivations for the research illustrated in this thesis emerge. Remarking that all the numbers reported in table 1.1 and table 1.2 are, in any way, finite values found estimating a quantity that is not finite under the considered prior settings, the numerical instability of the MCMC estimation outcomes are evident with small sample sizes. This instability manifests itself in two distinct ways: disproportionately large estimates are found for the considered posterior summaries, otherwise different values (often apparently reliable) for the same quantity are estimated in different runs of the algorithm. Unfortunately, these warnings vanish with the sample size increase and the user is led to believe in the estimates validity.

1.3 Work summary

As hinted before, the issues affecting the Bayesian estimation of the log-normal mean were faced by Fabrizi and Trivisano (2012) and Fabrizi and Trivisano (2016), wherein the log-normal linear model was considered. The core of their proposal consists of specifying a generalized inverse Gaussian (GIG) prior for the variance in the log-scale σ^2. In this way,

existence conditions for the posterior moments of the target functionals to estimate were found and a careful inferential procedure in the Bayesian framework was proposed.
The aim of this work is to fill the gap which is present in the literature approaching the estimation of log-normal quantiles (both unconditional and conditional) and of the log-normal linear mixed model form a Bayesian perspective, proposing a mathematically coherent inferential procedure.
The thesis is divided into two parts. In the first one, the quantile estimation problem is faced. In particular, in chapter 2, a new distribution is derived and its properties are reported, and in chapter 3 it is shown that this distribution is crucial in the posterior inference on the target functional θ_p and the Bayes estimators are derived; then, in chapter 4, they are first evaluated in a simulation study and then applied to real data.
The second part about the estimation of the log-normal mixed model is organized with a similar scheme: in chapter 5, the mathematical framework to solve the inferential problem is developed, whereas in chapter 6, a simulation study is carried out and the proposed methods are applied to real data.
Moreover, a particular care to computational aspects was taken in order to facilitate and encourage the practitioners to use the developed methods: an `R` package named `BayesLN` is implemented, whose manual is reported in appendix E. It is aimed at enclose functions useful in carrying out a proper Bayesian analysis under the log-normality assumption for data. The methods developed in Fabrizi and Trivisano (2012) and Fabrizi and Trivisano (2016) are implemented in the package too.

Part I

Inference on log-normal quantiles

Chapter 2

The SMNG Distribution

Before addressing the focus on the main topic of this part, that is the Bayesian estimation of log-normal quantiles, some preliminary results are required. The chapter is organized as follows: an introduction to the generalized inverse Gaussian (GIG) distribution is provided in section 2.1, whereas a new family of normal mean-variance mixtures is presented in section 2.2. In section 2.3, the new distribution with the GIG as mixing distribution is derived and its properties are studied, in section 2.4 it is compared to the generalized hyperbolic distribution and its exponential transformation is considered in section 2.5. Finally, some computational details are provided in section 2.6.

2.1 Generalized Inverse Gaussian Distribution

The generalized inverse Gaussian, a general positive real valued distribution, is characterized by the following probability density function:

$$f_X(x) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{1}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \gamma^2 x\right)\right\}, \quad x > 0; \tag{2.1}$$

where $K_\nu(x)$ is the Bessel K function (see appendix A.2). The three parameters $(\lambda, \delta, \gamma)$ could assume values within the following ranges:

- General case: $\lambda \in \mathbb{R},\ \delta > 0,\ \gamma > 0$;
- Limiting case I: $\lambda > 0,\ \delta \to 0,\ \gamma > 0$;
- Limiting case II: $\lambda < 0,\ \delta > 0,\ \gamma \to 0$.

By considering the relation (A.7), it is evident that the first limiting case corresponds to the *gamma distribution*, whereas the second one leads to the *inverse gamma distribution*. Other known distributions might be deduced with particular values of the parameters, like the *exponential distribution* and the *inverse Gaussian distribution* (Paolella, 2007).

On the other hand, the GIG distribution is strictly connected to the positive α-stable distribution. The α-stable distribution represents a very flexible 4-parameters distribution that is largely characterized by the so called stability parameter α, whose value controls the tail heaviness. Usually, the density of this family of distributions is not available in closed form and the law is defined in terms of characteristic function. The positive α-stable distribution is deduced when the skewness parameter assumes the value 1, implying the complete positive skewness. It is possible to prove that the GIG distribution can be obtained starting from the positive α-stable one by applying determined mathematical transforms aimed at obtaining lighter tailed distributions and by considering the case $\alpha = 1/2$ (Meyers, 2010).
It is possible to consider λ as the *shape* parameter, δ a *scale* parameter, whereas γ controls the tail heaviness. Figure 2.1 gives an idea of the GIG behaviour with respect to different parameter values. In particular, the role of γ is emphasized by plotting the log-density function: it is clear that the lower the value of γ is, the tail of the distribution is heavier. The GIG distribution has all the moments defined for $\gamma > 0$, because of its exponential decay in the tail, and they assume the form:

$$\mathbb{E}\left[X^j\right] = \left(\frac{\delta}{\gamma}\right)^j \frac{K_{\lambda+j}(\delta\gamma)}{K_\lambda(\delta\gamma)}. \tag{2.2}$$

Besides, the moment generating function for the general case is:

$$\mathbb{M}_X(r) = \left(\frac{\gamma}{\sqrt{\gamma^2 - 2r}}\right)^\lambda \frac{K_\lambda(\delta\sqrt{\gamma^2 - 2r})}{K_\lambda(\delta\gamma)}, \quad r < \frac{\gamma^2}{2}. \tag{2.3}$$

Another useful quantity to report in order to characterize the distribution is the mode:

$$Mo(X) = \frac{(\lambda - 1) + \sqrt{(\lambda - 1)^2 + \delta^2\gamma^2}}{\gamma^2}; \tag{2.4}$$

which is particularly appealing since it is the unique synthetic measure of the GIG distribution that is free of Bessel K functions, and hence it is easily tractable. A detailed analysis of the distribution properties could be found in Jorgensen (1982).

2.1.1 The Extended GIG distribution

The power transformation of a GIG random variable is of interest. Let us consider a random variable $X \sim GIG(\lambda, \delta, \gamma)$, then the transformed variable $Y = X^{\frac{1}{\theta}}$ is distributed as an extended GIG (EGIG) distribution with parameters $(\lambda, \delta, \gamma, \theta)$. The properties of this distribution were studied in Silva et al. (2006): its probability density function can be easily obtained by applying the random variable transformation formula, and the moments immediately follow from the (2.2) with $\mathbb{E}\left[Y^j\right] = \mathbb{E}\left[X^{\frac{j}{\theta}}\right]$.

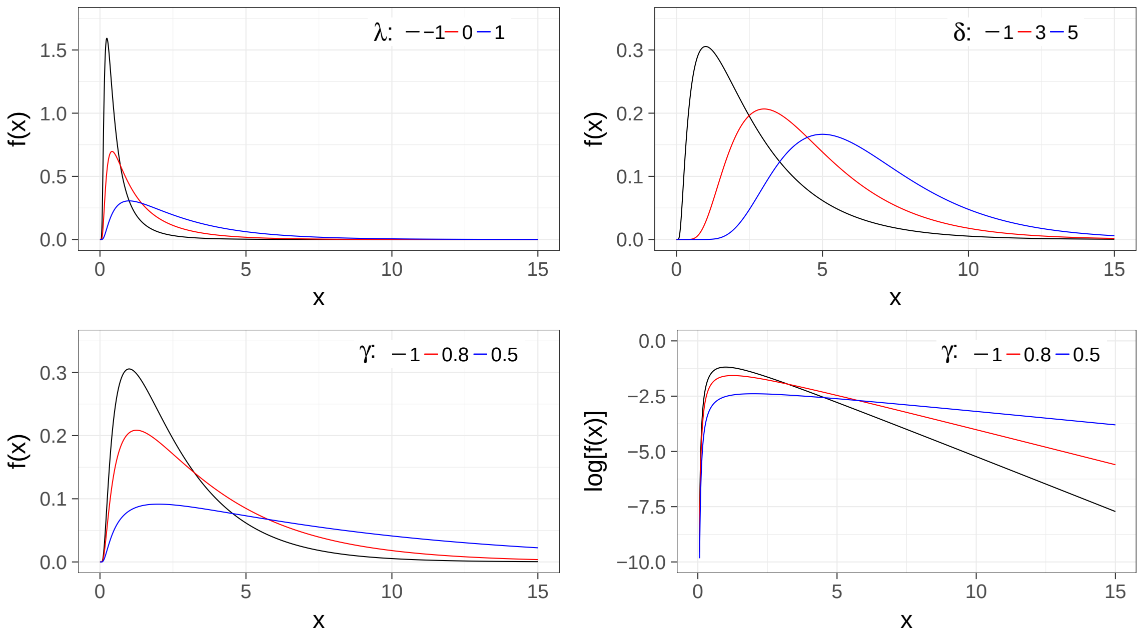

Figure 2.1: Plots of the GIG density function with different values of the parameters λ, δ and γ. In each plot the other parameters are fixed equal to 1. To point out the impact of γ on the tail the log-density is also reported for that parameter.

2.2 Just another normal mean-variance mixture

A widespread and flexible family of distributions, which is employed to model data endowed with particular features (e.g. multi-modality, heterogeneity, marked skewness and heavy tails) that cannot usually be captured by standard density functions, is the *normal mean-variance mixture*. The main properties of these distributions were firstly studied in Barndorff-Nielsen et al. (1982), but it is still an open research field (Yu, 2017). A fundamental paper, that is considered to be the starting point of the idea of mixture distribution, is the one by Barndorff-Nielsen (1977), where the author derived the generalized hyperbolic (GH) distribution to model the wind blown sand particle size (see appendix B.1).
In general, the univariate probability distributions belonging to this family are obtained choosing a *mixing density* $g(\cdot)$ for the random variable W, that is a non-negative real-valued distribution, and defining the random variable:

$$X = \mu + \beta W + \sqrt{W} Z, \tag{2.5}$$

where $\mu, \beta \in \mathbb{R}$ are constants. Furthermore, Z is distributed as a standard normal and it is independent of the mixing variable W. The aim of a random variable defined in this way is to gain flexibility in modelling by introducing variability both in the mean and in the variance of a Gaussian distribution.

From this general definition, it is possible to obtain several mixture distributions by considering different *mixing distributions*. In fact, the GH case cited above is characterized by the assumption that W is distributed as a GIG. On the other hand, also the normality assumption, executed by setting $Z \sim \mathcal{N}(0,1)$, might be relaxed employing other distributions like the skew normal (Arslan, 2015).
In this section, a new family of distributions strictly related to the normal mean-variance mixture is introduced. It could be named *scale-mean mixture of normal distribution* and, to my knowledge, it has not received any attention in the literature.

Definition 2.1. *Considering two independent distributions $Z \sim \mathcal{N}(0,1)$ and $W \sim g(\cdot)$, non-negative real valued random variable, then the distribution defined as:*

$$X = \mu + \beta\sqrt{W} + \sqrt{W}Z, \tag{2.6}$$

$\mu, \beta \in \mathbb{R}$, is a scale-mean mixture of normal distribution.

A random variable defined in this way, has the following conditional distribution:

$$X|W = w \sim \mathcal{N}\left(\mu + \beta\sqrt{w}, w\right); \tag{2.7}$$

and it is similar to the standard Gaussian mean-variance mixture, since it simply introduces a lower degree of variability in the mean. In fact, the term $\sqrt{W}$ appear in the (2.7) instead of W, which characterizes the usual normal mean-variance mixture.

2.3 A new mixture with GIG as mixing distribution

In this section, the *scale-mean mixture of normal distribution* is considered in the case $W \sim GIG$. A parallel result to the one related to the GH distribution is deduced and the derived distribution is labelled as SMNG, which synthesizes *a Scale-Mean mixture of Normal distribution assuming a GIG distribution on the scale.*

Theorem 2.1 (SMNG Distribution)**.** *Let consider the following scale-mean mixture of normal distribution:*

$$X|W = w \sim \mathcal{N}\left(\mu + \beta\sqrt{w}, w\right), \qquad W \sim GIG(\lambda, \delta, \gamma); \tag{2.8}$$

then the random variable X marginally assumes a SMNG distribution with parameters $(\lambda, \delta, \gamma, \mu, \beta)$, under the conditions $\delta, \gamma > 0$. Besides, the probability density function of X can be expressed:

(i) in integral form:

$$\begin{aligned} f_X(x) =& c(\lambda,\delta,\gamma,\beta) \int_0^\infty t^{-2\lambda} \exp\left\{ -\frac{1}{2}\Bigg([(x-\mu)^2 + \delta^2]t^2 + \right. \\ & \left. + \frac{\gamma^2}{t^2} - 2\beta(x-\mu)t \Bigg) \right\} \mathrm{d}t; \end{aligned} \tag{2.9}$$

(ii) as an infinite sum:

$$f_X(x) = c(\lambda,\delta,\gamma,\beta)\sum_{i=0}^{+\infty}\frac{[\beta(x-\mu)]^i}{i!}K_{\lambda-\frac{i+1}{2}}\left(\gamma\sqrt{\delta^2+(x-\mu)^2}\right)\times \left(\frac{\gamma}{\sqrt{\delta^2+(x-\mu)^2}}\right)^{\lambda-\frac{i+1}{2}}; \tag{2.10}$$

where:

$$c(\lambda,\delta,\gamma,\beta) = \frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)}e^{-\frac{\beta^2}{2}}. \tag{2.11}$$

Proof. (i) To find the marginal density function of X, the following integral needs to be solved:

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{\mathcal{N}}(x|w)f_{GIG}(w)\mathrm{d}w \\ &= \frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{2\sqrt{2\pi}K_\lambda(\delta\gamma)}\int_0^\infty w^{\lambda-\frac{1}{2}-1}\exp\left\{-\frac{1}{2}\left(\frac{(x-\mu-\beta\sqrt{w})^2}{w}+ \right.\right. \\ &\qquad \left.\left. +\frac{\delta^2}{w}+\gamma^2 w\right)\right\}\mathrm{d}w. \end{aligned} \tag{2.12}$$

By using simple algebra it can be noted that:

$$\frac{(x-\mu-\beta\sqrt{w})^2}{w} = \frac{(x-\mu)^2}{w} - \frac{2\beta(x-\mu)}{\sqrt{w}} + \beta^2. \tag{2.13}$$

If this expression is plugged into the integral and the change of variable $w = t^{-2}$ is performed ($\mathrm{d}w = -2t^{-3}\mathrm{d}t$), the result of equation (2.9) is deduced.

(ii) To obtain the infinite sum, it is required to recognize that the integral in (2.9) is the kernel of the Laplace transformation of the EGIG density function with parameters $\left(-\lambda+\frac{1}{2},\sqrt{\delta^2+(x-\mu)^2},\gamma,2\right)$, whose moments are known (see section 2.1.1) and can be used as follow:

$$f_X(x) = \frac{\left(\frac{\gamma}{\delta}\right)^\lambda e^{-\frac{\beta^2}{2}}}{\sqrt{2\pi}K_\lambda(\delta\gamma)}\frac{K_{-\lambda+\frac{1}{2}}(\gamma\sqrt{\delta^2+(x-\mu)^2})}{(\sqrt{\delta^2+(x-\mu)^2}/\gamma)^{-\lambda+\frac{1}{2}}}\int_0^\infty e^{rt}f_{EGIG}(t)\mathrm{d}t, \tag{2.14}$$

where $r = \beta(x-\mu)$.

To complete the proof and get the (2.10) it is necessary to expand the exponential inside the integral. Since the series and the integral are bounded, the summation and the integral can be exchanged:

$$\int_0^\infty e^{rt}f_{EGIG}(t)\mathrm{d}t = \sum_{i=0}^{+\infty}\frac{r^i}{i!}\int_0^\infty t^i f_{EGIG}(t)\mathrm{d}t. \tag{2.15}$$

It is possible to recognize that the obtained integrals inside the sum are the moments of order i of the EGIG distribution. By substituting their expressions the final result is obtained. ■

Unfortunately, it is not possible to obtain a representation of the SMNG density function without an integral or an infinite sum. In fact, the integral in (2.9) cannot be expressed in term of known special functions and it is not possible to apply the formula (A.10), i.e. the *multiplication theorem* of the Bessel K function to the infinite sum in (2.10) because of the fractional index in the order.
Even if the convergence of the series (2.10) is a consequence of the equivalence with the convergent integral (2.9), it is also possible to prove it analytically. In fact, considering the standard ratio test, the equivalence (A.5) and the approximation (A.8), the generic term of the sum with $j \to +\infty$ and $\nu = -\lambda + \frac{j+1}{2}$ is:

$$a_j = \frac{[\beta(x-\mu)]^j}{j!}\sqrt{\frac{2}{\pi}}\left(\frac{e\left[\delta^2+(x-\mu)^2\right]}{2}\right)^{\nu}\nu^{-\nu-\frac{1}{2}}. \tag{2.16}$$

Consequently, the ratio test assumes the form:

$$\frac{a_{j+1}}{a_j} \to \frac{1}{j\sqrt{\frac{j}{2}+1-\lambda}}\left(\frac{-\lambda+\frac{j+1}{2}}{-\lambda+\frac{j}{2}+1}\right)^{\frac{j}{2}+1-\lambda} = 0, \quad j \to +\infty, \tag{2.17}$$

and the series is absolutely convergent.

2.3.1 Meaning of the parameters

The influence of the five parameters is really similar to the ones of the GH distribution. Figure 2.1 is aimed at giving an idea of the implications that different parameters values have on the SMNG density function. The parameter μ operates on *location* and it induces a shift for the density. Then, it is possible to say that γ inherits its role in the GIG distribution and it is a *shape* parameter: smaller values imply heavier tails, as it is possible to see observing the log-density plots. Its strict connection to the tail heaviness is also evident from its involvement in the moment generating function existence condition, as will be investigated in the following sections. Another parameter that influences the tails is λ. On the other hand, β is an *asymmetry* parameter, whereas δ is the *scale* parameter: with smaller values the distribution is more concentrated around the peak.
Another useful result about the parameters is the behaviour of the SMNG distribution with respect to changes in location and scale.

Proposition 2.1 (Location-scale behaviour)**.** *If* $X \sim SMNG(\lambda,\delta,\gamma,\mu,\beta)$ *and two constant* $a \in \mathbb{R}\setminus\{0\}$, $b \in \mathbb{R}$ *are considered, then:*

$$aX + b \sim SMNG\left(\lambda, |a|\delta, \frac{\gamma}{|a|}, \beta, a\mu + b\right). \tag{2.18}$$

Proof. To prove this result, it is sufficient to obtain the density function of the transformed random variable:

$$f_{aX+b}(x) = a^{-1} f_X\left(\frac{x-b}{a}; \lambda, \delta, \gamma, \beta, \mu\right). \tag{2.19}$$

After simple algebra and a change of variable into the integral, it is possible to get:

$$\begin{aligned} f_{aX+b}(x) =& \frac{\left(\frac{\gamma}{a^2\delta}\right)^{\lambda}}{\sqrt{2\pi} K_\lambda\left(a\delta a^{-1}\gamma\right)} e^{-\frac{\beta^2}{2}} \int_0^\infty z^{-2\lambda} \times \\ & \times \exp\left\{ -\frac{1}{2}\left(\left[(x-b-a\mu)^2 + a^2\delta^2\right] z^2 + \frac{\gamma^2}{a^2 z^2} + \right.\right. \\ & \left.\left. - 2\beta(x-b-a\mu)z \right)\right\} \mathrm{d}z \end{aligned} \tag{2.20}$$

Fixing $\tilde{\delta} = |a|\delta$, $\tilde{\gamma} = \frac{\gamma}{|a|}$ and $\tilde{\mu} = a\mu + b$ and comparing the previous density in (2.9) it is clear that it is again the density function of a SMNG distribution with parameters $\left(\lambda, \tilde{\delta}, \tilde{\gamma}, \beta, \tilde{\mu}\right)$. ■

2.3.2 Moment generating function and moments of the distribution

The general results that can be obtained for the conventional cases of normal mean-variance mixtures (Hammerstein, 2010) do not hold for the SMNG distribution because of the particular form of the density function. Therefore, it is necessary to algebraically deduce the quantities that characterize the distribution, beginning from the moment generating function.

Theorem 2.2 (SMNG Moment Generating Function)**.** *Considering a random variable $X \sim SMNG(\lambda, \delta, \gamma, \mu, \beta)$, it has a moment generating function of the form:*

$$\mathbb{M}_X(r) = e^{\mu r} \frac{\left(\frac{\gamma}{\sqrt{\gamma^2 - r^2}}\right)^{\lambda}}{K_\lambda(\delta\gamma)} \sum_{i=0}^{+\infty} \frac{(r\beta)^i}{i!} \left(\frac{\delta}{\sqrt{\gamma^2 - r^2}}\right)^{\frac{i}{2}} K_{\lambda + \frac{i}{2}}\left(\delta\sqrt{\gamma^2 - r^2}\right), \tag{2.21}$$

that is defined if $r < \gamma$.

Proof. To get simpler computations, the case $\mu = 0$ is considered. Recalling the integral form of the density function (2.9), using the definition of moment generating function:

$$\begin{aligned} \mathbb{M}_X(r) = \frac{\left(\frac{\gamma}{\delta}\right)^{\lambda}}{\sqrt{2\pi} K_\lambda(\delta\gamma)} e^{-\frac{\beta^2}{2}} \int_{-\infty}^{+\infty} e^{rx} \int_0^\infty t^{-2\lambda} \times \\ \times \exp\left\{ -\frac{1}{2}\left([x^2 + \delta^2] t^2 + \frac{\gamma^2}{t^2} - 2\beta x t \right)\right\} \mathrm{d}t\mathrm{d}x. \end{aligned} \tag{2.22}$$

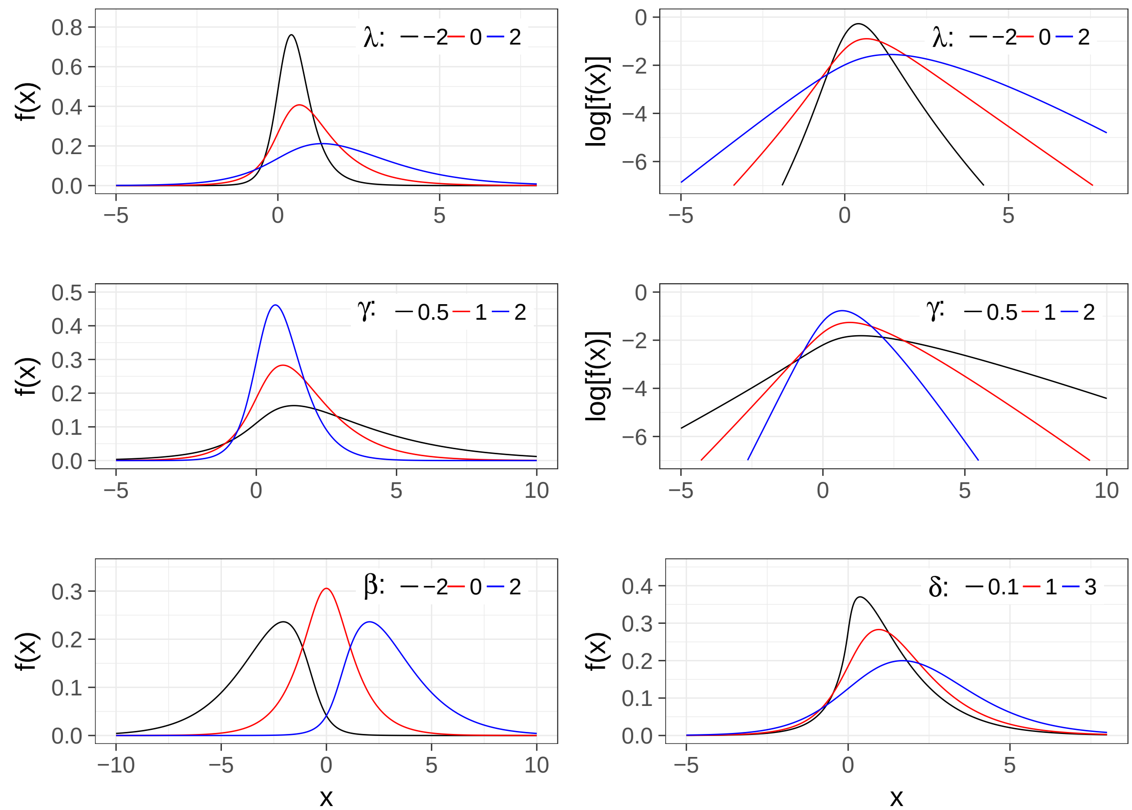

Figure 2.2: Comparison of the SMNG density with different values of the parameters. In the legend the varying parameters are showed, the other are fixed equal to 1 with the exception of $\mu = 0$. For λ and γ also the logarithm of the density is reported.

By applying Fubini's theorem and after a change of variable it is possible to recognize the integral of the Gaussian distribution moment generating function. It can be solved using the formula 3.323.2 in Gradshteyn and Ryzhik (2014):

$$\begin{aligned}\mathbb{M}_X(r) &= \frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)} e^{-\frac{\beta^2}{2}} \int_0^\infty t^{-2\lambda-1} \exp\left\{-\frac{1}{2}\left(\delta^2 t^2 + \frac{\gamma^2}{t^2}\right)\right\} \times \\ &\quad \times \int_{-\infty}^{+\infty} \exp\left\{-\frac{z^2}{2} + \frac{z(r+\beta t)}{t}\right\} \mathrm{d}z\mathrm{d}t = \\ &= \frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{K_\lambda(\delta\gamma)} \int_0^\infty t^{-2\lambda-1} \exp\left\{-\frac{1}{2}\left(\delta^2 t^2 + \frac{\gamma^2 - r^2}{t^2} - \frac{2r\beta}{t}\right)\right\} \mathrm{d}t.\end{aligned} \tag{2.23}$$

The latter integral is convergent if $r < \gamma$ and, through another change of variable, an integral with the same structure as the one in (2.14) is obtained. Therefore, if the same procedure used to prove the statement *(ii)* of Theorem 2.1 is applied, the (2.21) can be deduced (up to the term $e^{\mu r}$).

Finally, the result can be extended to any μ by applying Proposition 2.1 and the formula of the moment generating function of a linearly transformed random variable. ■

In order to derive a generic expression for the SMNG distribution moments, it is useful to start from the particular case $\mu = 0$.

Proposition 2.2 (SMNG j-th Moment). *If $Z \sim SMNG(\lambda, \delta, \gamma, \beta, 0)$, then the j-th moment assumes the following form:*

$$\mathbb{E}\left[Z^j\right] = \begin{cases} \left(2\frac{\delta}{\gamma}\right)^{\frac{j}{2}} \frac{K_{\lambda+\frac{j}{2}}(\delta\gamma)}{K_\lambda(\delta\gamma)} \frac{\Gamma\left(\frac{j+1}{2}\right)}{\sqrt{\pi}} \Phi\left(-\frac{j}{2}, \frac{1}{2}; -\frac{\beta^2}{2}\right) & j \text{ even} \\ \beta\left(2\frac{\delta}{\gamma}\right)^{\frac{j}{2}} \frac{K_{\lambda+\frac{j}{2}}(\delta\gamma)}{K_\lambda(\delta\gamma)} \frac{\sqrt{2}\Gamma\left(\frac{j}{2}+1\right)}{\sqrt{\pi}} \Phi\left(\frac{1-j}{2}, \frac{3}{2}; -\frac{\beta^2}{2}\right) & j \text{ odd} \end{cases} \tag{2.24}$$

where $\Phi(\cdot,\cdot;\cdot)$ is the Kummer's M confluent hypergeometric function (see appendix A.3).

Proof. The j-th moment of Z is defined as:

$$\begin{aligned} \mathbb{E}\left[Z^j\right] =& \frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)} \int_{-\infty}^{+\infty} z^p \int_0^{+\infty} t^{-2\lambda} \exp\Bigg\{-\frac{1}{2}\Bigg[(zt-\beta)^2 + \\ &+ \frac{\gamma^2}{t^2} + \delta^2 t^2\Bigg]\Bigg\} \mathrm{d}x\mathrm{d}t = \end{aligned} \tag{2.25}$$

$$\begin{aligned} =& \frac{\left(\frac{\gamma}{\delta}\right)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)} \int_0^{+\infty} t^{-2\lambda} \exp\left\{-\frac{1}{2}\left[\frac{\gamma^2}{t^2} + \delta^2 t^2\right]\right\} \times \\ &\times \left(\int_{-\infty}^{+\infty} z^j \exp\left\{-\frac{1}{2}\left[(zt-\beta)^2\right]\right\} \mathrm{d}z\right) \mathrm{d}t; \end{aligned} \tag{2.26}$$

by Fubini's theorem. With the substitution $zt = y$, in the inner integral the j moment of a $\mathcal{N}(\beta, 1)$ can be recognized.
Recalling that the j-th moment of a Gaussian distribution with mean μ and variance σ^2 is defined as (Winkelbauer, 2012):

$$\mathbb{E}[Y^j] = (i\sigma)^j \exp\left\{-\frac{\mu^2}{4\sigma^2}\right\} D_j\left(-i\frac{\mu}{\sigma}\right), \tag{2.27}$$

where $D_\nu(x)$ is the Parabolic Cylinder function (see appendix A.4), then applying the integral representation (A.4) to the (2.26) it is obtained:

$$\mathbb{E}\left[Z^j\right] = \left(\frac{\delta}{\gamma}\right)^{\frac{j}{2}} \frac{K_{\lambda+\frac{j}{2}}(\delta\gamma)}{K_\lambda(\delta\gamma)} (i)^j \exp\left\{-\frac{\beta^2}{4}\right\} D_j(-i\beta). \tag{2.28}$$

By using the expression of $D_\nu(x)$ as a function of the Kummer's M function (A.15) and by applying the Euler reflection formula (A.2) to the gamma functions the final form (2.24) is reached. ■

Then, if a generic SMNG distribution is considered, through Proposition 2.2 it is possible to get an expression for the moments of the form $\mathbb{E}\left[(X-\mu)^j\right]$. However, the following lemma gives the possibility to use this particular result in order to compute both the raw moments $\mathbb{E}\left[X^j\right]$ and the central moments $\mathbb{E}\left[(X-\mathbb{E}[X])^j\right]$ of any SMNG distribution.

Lemma 2.1. *For any constant a and b and any positive integer j, it is true that:*

$$\mathbb{E}\left[(X-b)^j\right] = \sum_{l=0}^{j} \binom{j}{l} (a-b)^{j-l} \mathbb{E}\left[(X-a)^l\right]. \tag{2.29}$$

Proof. See Scott et al. (2011). ■

After the general definition of the moments and of the moment generating function, the characterization of the distribution can be completed writing the expected value:

$$\mathbb{E}[X] = \mu + \beta \left(\frac{\delta}{\gamma}\right)^{\frac{1}{2}} \frac{K_{\lambda+\frac{1}{2}}(\delta\gamma)}{K_\lambda(\delta\gamma)}; \tag{2.30}$$

and the variance:

$$\begin{aligned}\mathbb{V}[X] =& \beta^2 \left(\frac{\delta}{\gamma}\right) \left(\frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \frac{K^2_{\lambda+\frac{1}{2}}(\delta\gamma)}{K^2_\lambda(\delta\gamma)}\right) + \\ &+ \left(\frac{\delta}{\gamma}\right) \frac{K_{\lambda-1}(\delta\gamma) + K_{\lambda+1}(\delta\gamma)}{2K_\lambda(\delta\gamma)} - \frac{\lambda}{\gamma^2}.\end{aligned} \tag{2.31}$$

These expressions could be deduced by applying the $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$ operators to the formulation in (2.6) or through the moment generating function.

2.3.3 Particular cases

Since the GIG distribution includes two important limiting cases, which coincide with the inverse gamma and the gamma distributions, it is interesting to explore the resultant mixture distributions.

Inverse Gamma as mixing distribution

If the GIG distribution has $\lambda < 0$ and $\gamma \to 0$, then the random variable is an inverse gamma $W \sim IG(\alpha, \theta)$:

$$f_W(w) = \frac{\theta^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} e^{-\theta w^{-1}}, \qquad w > 0. \tag{2.32}$$

In this case, the integral (2.9) simplifies and it is possible to recognize the integral form of the parabolic cylinder function (A.14). Therefore, the marginal density function of the

SMNG reduces to:

$$f_X(x) = \frac{\theta^\alpha \Gamma(2\alpha+1) D_{-2\alpha-1}\left(\frac{\beta(\mu-x)}{\sqrt{(x-\mu)^2+2\theta}}\right)}{\sqrt{2\pi}\Gamma(\alpha)\left[(x-\mu)^2+2\theta\right]^{\alpha+\frac{1}{2}}} \times \\ \times \exp\left\{\frac{\beta^2(\mu-x)^2}{4\left[(x-\mu)^2+2\theta\right]} - \frac{\beta^2}{2}\right\}, \tag{2.33}$$

where $\theta = \frac{\delta^2}{2}$ and $\alpha = -\lambda$ if the parametrization of (2.8) is considered.
In agreement with the existence condition of the SMNG moment generating function (see theorem 2.2), it turns out that in this particular case $\mathbb{M}_X(r)$ is not defined.

Gamma as mixing distribution

If the limiting case I of the GIG distribution is examined, i.e. $\lambda > 0$ and $\delta \to 0$, the gamma distribution of parameters λ and $\nu = \frac{\gamma^2}{2}$ is the mixing distribution and the density of X is:

$$f_X(x) = \frac{\sqrt{2}\nu^\lambda}{\sqrt{\pi}\Gamma(\lambda)} e^{-\frac{\beta^2}{2}} \sum_{i=0}^{+\infty} \frac{\left[\beta(x-\mu)\right]^i}{i!} K_{\lambda-\frac{i+1}{2}}\left(\sqrt{2\nu}|x-\mu|\right) \times \\ \left(\frac{\sqrt{2\nu}}{|x-\mu|}\right)^{\lambda-\frac{i+1}{2}}. \tag{2.34}$$

It has the same structure of the (2.10) and the existence conditions of the moment generating functions remain the same as in the general case.

2.4 Comparison with the GH distribution

The distribution considered in Theorem 2.1 defines a real-valued random variable that has a similar behaviour with respect to the GH distribution. First of all, if it is fixed $\beta = 0$ the SMNG distribution assumes the same limiting case of the GH, i.e. the symmetric GH distribution.
Since the two distributions depend on 5 parameters that have the same meaning, it is interesting to observe the changes in the densities taking equal parameters sets. In figure 2.3, the main difference between the two distributions clearly appears: the right tail (left in case of negative β) is considerably lighter for the SMNG distribution.
In effect, the density of a GH distribution in the tails is:

$$f_{GH}(x) = c|x|^{\lambda-1} \exp\left\{-\sqrt{\gamma^2+\beta^2}|x| + \beta x\right\}, \qquad |x| \to \infty \tag{2.35}$$

that defines a distribution with *semi-heavy tails*; whereas the tails of the SMNG distribution, recalling the (A.9), have the following law:

$$f_{SMNG}(x) = c|x|^{\lambda-1} \exp\left\{-\gamma|x| + \sqrt{\gamma}\beta|x|^{\frac{1}{2}}\right\}. \tag{2.36}$$

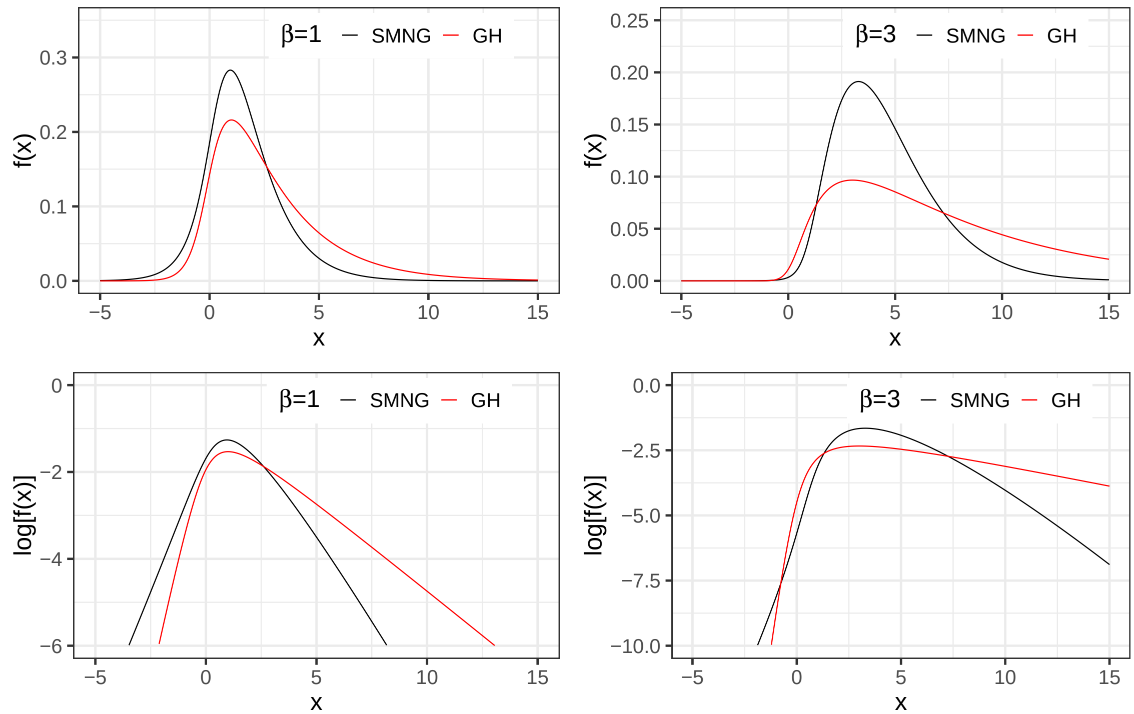

Figure 2.3: Plots of GH and SMNG densities and log-densities with the same parameters fixed: the cases $\beta = 1$ and $\beta = 3$ are reported. The other parameters are fixed equal to 1 with the exception of $\mu = 0$.

Therefore, for each $\beta > 0$ and it is possible to conclude that:

$$f_{GH}(x_0) > f_{SMNG}(x_0), \qquad \forall x_0 > M, \tag{2.37}$$

where M is a large positive number. The same holds for the left tail with $\beta < 0$ and M negative. On the other hand, if the left tail with $\beta > 0$ is considered, then the GH density decays faster than the SMNG density.

2.5 The log-SMNG distribution

In the subsequent parts of this work, the distribution of the exponential transformation of a SMNG distributed random variable is of interest. This kind of distribution can be defined as follow.

Definition 2.2 (Log-SMNG distribution)**.** *If $X \sim SMNG(\lambda, \delta, \gamma, \beta, \mu)$, then the random variable $Y = \exp\{X\}$ assumes a log-SMNG distribution. Equivalently, the log-SMNG distribution might be defined as the random variable whose logarithmic transformation is distributed as a SMNG.*

Therefore, a continuous distribution that assumes positive values only is faced. With a simple application of the random variable transformation formula, it is possible to deduce the key properties that characterize the distribution.

Proposition 2.3 (Log-SMNG characterization)**.** *If* $X \sim SMNG(\lambda, \delta, \gamma, \beta, \mu)$ *and* $Y = \exp\{X\}$*, then* Y *possesses:*

(i) probability density function:

$$f_Y(y) = \frac{1}{y} f_X\left(\log[y]\right), \tag{2.38}$$

where $f_X(\cdot)$ *is defined in the* (2.9) *or* (2.10)*;*

(ii) expectation and j-th central moment (defined if $\gamma > j$*):*

$$\mathbb{E}[Y] = e^{\mu} \frac{\left(\frac{\gamma}{\sqrt{\gamma^2-1}}\right)^{\lambda}}{K_\lambda(\delta\gamma)} \sum_{i=0}^{+\infty} \frac{(\beta)^i}{i!} \left(\frac{\delta}{\sqrt{\gamma^2-1}}\right)^{\frac{i}{2}} K_{\lambda+\frac{i}{2}}\left(\delta\sqrt{\gamma^2-1}\right), \tag{2.39}$$

$$\mathbb{E}[Y^j] = e^{j\mu} \frac{\left(\frac{\gamma}{\sqrt{\gamma^2-j^2}}\right)^{\lambda}}{K_\lambda(\delta\gamma)} \sum_{i=0}^{+\infty} \frac{(j\beta)^i}{i!} \left(\frac{\delta}{\sqrt{\gamma^2-j^2}}\right)^{\frac{i}{2}} K_{\lambda+\frac{i}{2}}\left(\delta\sqrt{\gamma^2-j^2}\right). \tag{2.40}$$

Proof. (i) The result follows from the simple application of the random variable transformation formula.
(ii) Given that $Y = \exp\{X\}$, then $\mathbb{E}[Y^j] = \mathbb{E}[\exp\{jX\}]$. It is the moment generating function of the SMNG distribution defined in the (2.21) evaluated in j. Consequently, the existence of the moments of the log-SMNG is regulated by the same existence condition of the SMNG moment generating function. ■

2.6 Computational notes and software implementation

In order to use the two distributions described in this chapter, `R` (R Core Team, 2017) functions needed to be implemented. In particular, to generate random samples from the SMNG distribution, the mixture specification in (2.8) was exploited. Therefore, to obtain an independent sample of size n the following steps were executed:

- generate a sample of n independent realizations from a GIG distribution using the function `rgig()`, included in the `R` package `ghyp` (Breymann and Lüthi, 2013);
- use `rnorm()` to generate from the resultant normal distribution.

Another function implements the SMNG density function. This is a tricky task because of the numerical instability of the two formulations given in Theorem 2.1. In fact, with high

values of the parameters δ, γ, β, numerical problems might be detected both for the function based on `integrate()` and the infinite sum, that includes a ratio of Bessel K functions. To overcome this issue, the best solution is to implement the sum in (2.10) with the option `expon.scaled=TRUE` of the function `besselK()`, that returns $\exp\{x\}K_\nu(x)$. In this way, the eventual problems related to the numerical underflow caused by the ratio of two Bessel K functions with elevate arguments can be avoided. Besides, an easy way to implement a stopping rule for the infinite sum consists in putting a test in order to evaluate the magnitude of each term with respect to the partial sum. However, a function that produces the density through the integral representation was implemented too.
Then, to get the cumulative distribution function:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t, \tag{2.41}$$

a numerical integration procedure was employed through the `integrate()` function.
Finally, the computation of the quantiles represents a famous problem of numerical inversion. In fact, the quantile α that corresponds to a probability p is the solution of the non-linear equation:

$$F_X(\alpha) - p = 0. \tag{2.42}$$

To solve it, the standard `uniroot()` procedure was employed (Lange, 2010).
The same ideas were used to implement the key functions related to the log-SMNG distribution, by applying the simple exponential transformation to the generated sample and the formula (2.38) for the density.
All these functions are implemented in `R` and included in the developed package `BayesLN`. The standard `R` denominations are adopted: `dSMNG` and `dlSMNG` evaluate the SMNG and the log-SMNG density functions, `pSMNG` and `plSMNG` the cumulative functions, `qSMNG` and `qlSMNG` the quantiles and, finally, `rSMNG` and `rlSMNG` allow to generate random numbers form the distributions. Moreover, the SMNG moments are implemented in the function `SMNGmoment`, whereas the function that evaluates the moment generating function is `SMNG_MGF`.

Chapter 3

Bayesian inference for the log-normal quantiles

The estimation of log-normal quantiles can be of interest in many applications. For example, in environmental monitoring and occupational health analyses, it is common to estimate extreme quantiles in the right tail of a skewed distribution from small samples (Bullock and Ignacio, 2006; Gibbons et al., 2009; Krishnamoorthy et al., 2011), or to compare a fixed legal exposure limit to an extreme quantile (or to its upper confidence limit, UCL) estimated from a sample that could be small. Under these conditions, the tools available in the current literature, that mainly consist of the exponentiation of standard frequentist results obtained in the log-scale, can produce unefficient point and interval estimators with poor coverage or low precision and can be significantly improved. The proposed methodology improves current methods, especially in the analysis of small samples. The estimation of quantiles is relevant in several other applied fields like the analysis of lifetime data (Lawless, 2003) or flood frequency analysis (Stedinger, 1980; Hamed and Rao, 1999).

The estimation of log-normal quantiles has received little attention so far. In the frequentist literature, Longford (2012) identifies a class of estimators depending on two constants that he determines with the aim of minimizing the frequentist mean square error (MSE); he overlooks relevant inferential problems such as interval estimation.

In this chapter the problem of Bayesian estimation of log-normal quantile is studied with a particular focus on the posterior moment finiteness. In section 3.1 an overview on the already developed methods for the log-normal quantile estimation is provided. All the mathematical details to obtain a rigorous Bayesian inferential framework are presented in section 3.2, while in section 3.3 the hyperparameter specification strategy is discussed. Finally, the methodology is extended to the conditional quantile estimation problem in section 3.4.

3.1 Estimation of log-normal quantiles: current state of the art

3.1.1 Non-parametric estimation

A non-parametric approach for the quantile estimation is often adopted. Even if the cornerstone on this work is the log-normality distributional assumption, it is worth to consider this estimation procedure that is commonly used also with small sample sizes.
There exist a lot of different methods easily accessible in common statistical software and a good review is the paper by Hyndman and Fan (1996). As a benchmark for the developed proposals, the standard `R` function `quantile` is used, with the default type 7 method. It is due to Gumbel (1939) and it is based on the empirical cumulative distribution function built on the following definition of k-th plotting position:

$$p_k = \frac{k-1}{n-1}, \tag{3.1}$$

that coincides with the mode of the distribution function of the k-th order statistic: $F(X_{(k)})$. Then, the quantile p, included between the positions $k-1$ and k, is obtained by linear interpolation:

$$\hat{Q}_p^7 = X_{(k-1)} - \left(X_{(k-1)} - X_{(k)}\right) \frac{p_{k+1} - p}{p_{k+1} - p_k}. \tag{3.2}$$

3.1.2 Naive estimation

When an estimate of the log-normal p-th quantile is required, the usual procedure is to take the exponential of the well known normal quantile formula:

$$\hat{\theta}_p = \exp\left\{\hat{\xi} + \Phi^{-1}(p)\hat{\sigma}\right\}, \tag{3.3}$$

plugging the unbiased estimates of the mean and variance in the log-scale: $\hat{\xi}, \hat{\sigma}^2$. It is worth to highlight that, after the squared root transformation, $\hat{\sigma}$ is not an unbiased estimator of the population standard deviation in the log-scale. Besides, even if a monotone transformation (as the exponential is) does not change the order statistics like the quantiles, it might affect all the desirable properties that the estimator has in the original scale.
The same procedure is applied to compute the extremes of the confidence intervals (Gibbons et al., 2009). In the two sided case with the fixed confidence level $1-\alpha$ they are:

$$\left[\exp\left\{\hat{\xi} + t_{\left(\frac{\alpha}{2}, n-1, k_p\right)} \frac{\hat{\sigma}}{\sqrt{n}}\right\}; \exp\left\{\hat{\xi} + t_{\left(1-\frac{\alpha}{2}, n-1, k_p\right)} \frac{\hat{\sigma}}{\sqrt{n}}\right\}\right], \tag{3.4}$$

where $t_{\left(\frac{\alpha}{2}, n-1, k_p\right)}$ is the quantile $\frac{\alpha}{2}$ of a non-central Student's t distribution with $n-1$ degrees of freedom and a non-centrality parameter $k_p = \sqrt{n}\Phi^{-1}(p)$.

In many applications, the one-sided intervals are particularly useful, and they are called lower confidence limit (LCL) or upper confidence limit (UCL):

$$
\begin{aligned}
LCL_p &= \exp\left\{\hat{\xi} + t_{(\alpha,n-1,k_p)}\frac{\hat{\sigma}}{\sqrt{n}}\right\}, \\
UCL_p &= \exp\left\{\hat{\xi} + t_{(1-\alpha,n-1,k_p)}\frac{\hat{\sigma}}{\sqrt{n}}\right\}.
\end{aligned} \tag{3.5}
$$

3.1.3 Longford's minimum MSE estimator

In the paper by Longford (2012), the following statistic was proposed to estimate the quantiles θ_p of a sample of observations assumed log-normally distributed:

$$Q_p = \exp\left\{\hat{\xi} + b_p\hat{\sigma} + d_p\hat{\sigma}^2\right\}, \tag{3.6}$$

where $\hat{\xi}$ and $\hat{\sigma}^2$ are the unbiased estimators of ξ and σ^2.
The values of the parameters (b_p, d_p) were fixed by minimizing the MSE of the estimator using the Newton-Raphson algorithm. The procedure was implemented by the author substituting the sample quantity $\hat{\sigma}^2$ to the unknown σ^2.
It is important to point out that the proposed estimator has finite expectation when d_p is negative or when:

$$\sigma^2 < \frac{n-1}{2d_p}, \tag{3.7}$$

where n is the sample size. The same inequality divided by 2 determines the existence condition for the MSE. These conditions are not testable since the variance σ^2 is not known.

3.2 Bayes estimator of the log-normal quantiles

The goal is to make inference on the log-normal quantiles θ_p defined in the (1.5) from the observed sample of size n: $(y_1, \dots, y_n)$. The logarithmic transformations of the observations are $w_i = \log(y_i)$.
Therefore, in order to obtain the posterior distribution of θ_p, it is necessary to assume a prior distribution for the two parameters ξ and σ^2. The conjugate Normal-GIG (NGIG) prior distribution is chosen (Thabane and Haq, 1999). Therefore, a normal prior conditionally on σ^2 is assumed for ξ and a generalized inverse Gaussian distribution is specified for σ^2:

$$\xi|\sigma^2 \sim \mathcal{N}\left(\xi_0, \frac{\sigma^2}{n_0}\right), \tag{3.8}$$

$$\sigma^2 \sim GIG(\lambda, \delta, \gamma). \tag{3.9}$$

Through the specification of the set of hyperparameters $(\xi_0, n_0, \lambda, \delta, \gamma)$ it is possible to express a huge amount of different prior situations. It is worth to highlight the fact that the

marginal prior assumed on ξ is a symmetric generalized hyperbolic distribution (see section B.1).
The usual non-informative improper priors like Jeffrey's prior, are not considered in this case because a key point of the work is to obtain a posterior distribution with finite moments, in order to deduce a Bayes estimator for the estimand, as it will be pointed out later.
Furthermore, to fix the notation, from now on the target functional in the log-scale will be defined as:

$$\begin{aligned}\eta_p &= \log\theta_p \\ &= \xi + \Phi^{-1}(p)\sigma;\end{aligned} \tag{3.10}$$

and the following sample quantities will be employed:

$$\bar{w} = \frac{\sum_{i=1}^{n} w_i}{n}; \tag{3.11}$$

$$v^2 = \frac{\sum_{i=1}^{n}(w_i - \bar{w})^2}{n}. \tag{3.12}$$

Before deducing the posterior distribution of θ_p, a series of useful results conditioned on σ^2 could be stated. They easily follow from the conjugacy of the NGIG prior with the Normal distribution.

Proposition 3.1. *If the Normal prior on ξ (3.8) is specified and the log-normal model is assumed for data, then the following results conditioned on σ^2 hold:*

(i)

$$\xi|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\xi_1, \frac{\sigma^2}{n_1}\right), \tag{3.13}$$

where $\xi_1 = \psi\bar{w} + (1-\psi)\xi_0$, $n_1 = n + n_0$ and $\psi = \frac{n}{n_1}$;

(ii)

$$\eta_p|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\bar{\eta}_p, \frac{\sigma^2}{n_1}\right), \tag{3.14}$$

where $\bar{\eta}_p = \xi_1 + \Phi^{-1}(p)\sigma$.

Proof. *(i)* It is necessary to consider:

$$\begin{aligned}p(\xi|\sigma^2, \mathbf{w}) &\propto p(\mathbf{w}|\xi, \sigma^2)p(\xi, \sigma^2) \\ &\propto \exp\left\{-\frac{\sum_{i=1}^{n}(w_i - \xi)^2}{2\sigma^2} - \frac{n_0(\xi - \xi_0)^2}{2\sigma^2}\right\} \\ &= \exp\left\{-\frac{\sum_{i=1}^{n}(w_i - \bar{w})^2}{2\sigma^2} - \frac{n_1(\psi\bar{w} + (1-\psi)\xi_0 - \xi)^2}{2\sigma^2}\right\} \\ &\propto \exp\left\{-\frac{n_1(\xi_1 - \xi)^2}{2\sigma^2}\right\};\end{aligned} \tag{3.15}$$

that is the kernel of a Gaussian distribution with mean ξ_1 and variance $\frac{\sigma^2}{n_1}$.
(ii) To prove this point it is possible to observe from (3.10) that η_p is a simple linear transformation of ξ and the same transformation can be applied to the normal distribution.

■

As a direct consequence of the previous proposition, the target functional θ_p assumes a log-normal distribution conditionally on σ^2.
Another important step is the derivation of the marginal posterior distributions for the parameters ξ and σ^2. Also in this case a standard result is faced, because of the conjugacy of the Normal prior on the Gaussian mean.

Proposition 3.2. *If the NGIG prior* (3.8), (3.9) *is specified and the log-normal model is assumed for data, then the posterior marginal distributions for the parameters* ξ *and* σ^2 *are:*

(i)

$$\xi|\mathbf{w} \sim GH\left(\bar{\lambda}, \bar{\gamma}, 0, \bar{\delta}, \xi_1\right), \tag{3.16}$$

where $\bar{\lambda} = \lambda - \frac{n}{2}$, $\bar{\gamma} = \sqrt{n_1}\gamma$ *and* $\bar{\delta} = \left(\sqrt{n_1}\right)^{-1}\sqrt{\delta^2 + nv^2}$.

(ii)

$$\sigma^2|\mathbf{w} \sim GIG\left(\bar{\lambda}, \bar{\delta}\sqrt{n_1}, \gamma\right). \tag{3.17}$$

Proof. *(i)* Integrating out σ^2 from the joint distribution and then applying (A.4):

$$\begin{aligned} p\left(\xi|\mathbf{w}\right) &\propto \int_0^{+\infty} p(\mathbf{w}|\xi,\sigma^2)p(\xi,\sigma^2)\mathrm{d}\sigma^2 \\ &\propto \int_0^{+\infty} (\sigma^2)^{-\frac{n+1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(nv^2 + n_1\left(\psi\bar{w} + (1-\psi)\xi_0 - \xi\right)^2\right)\right\} \times \\ &\qquad \times \left(\sigma^2\right)^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{\sigma^2} + \gamma^2\sigma^2\right)\right\} \mathrm{d}\sigma^2 \\ &= \int_0^{+\infty} (\sigma^2)^{\lambda-\frac{n+1}{2}-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2 + nv^2 + n_1(\xi-\xi_1)^2}{\sigma^2} + \gamma^2\sigma^2\right)\right\} \mathrm{d}\sigma^2 \\ &\propto \frac{K_{\lambda-\frac{n+1}{2}}\left(\gamma\sqrt{\delta^2 + nv^2 + n_1(\xi-\xi_1)^2}\right)}{\left(\sqrt{\delta^2 + nv^2 + n_1(\xi-\xi_1)^2}/\gamma\right)^{\frac{n+1}{2}-\lambda}}. \end{aligned} \tag{3.18}$$

The obtained expression is the kernel of a symmetric GH distribution (i.e. $\beta = 0$).
(ii) Starting from the same point of part *(i)* but integrating out ξ:

$$\begin{aligned} p\left(\sigma^2|\mathbf{w}\right) &\propto \int_{-\infty}^{+\infty} p(\mathbf{w}|\xi,\sigma^2)p(\xi,\sigma^2)\mathrm{d}\xi \\ &\propto (\sigma^2)^{\lambda-\frac{n+1}{2}-1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{\sigma^2} + \gamma^2\sigma^2\right)\right\} \times \\ &\qquad \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2\sigma^2}\left(n_1\left(\psi\bar{w} + (1-\psi)\xi_0 - \xi\right)^2\right)\right\} \mathrm{d}\xi. \end{aligned} \tag{3.19}$$

The integral is trivial since the kernel of a Gaussian distribution can be noted. Finally, the kernel of a GIG distribution can be recognized. ■

The main result of this section consists in the statement of the posterior distribution for θ_p.

Theorem 3.1. *If the NGIG prior* (3.8), (3.9) *is specified and the log-normal model is assumed for data, then:*

(i)

$$\eta_p|\mathbf{w} \sim SMNG(\bar{\lambda}, \bar{\delta}, \bar{\gamma}, \bar{\beta}, \xi_1), \tag{3.20}$$

where $\bar{\beta} = \sqrt{n_1}\Phi^{-1}(p)$;

(ii)

$$\theta_p|\mathbf{w} \sim \log SMNG(\bar{\lambda}, \bar{\delta}, \bar{\gamma}, \bar{\beta}, \xi_1). \tag{3.21}$$

Proof. *(i)* Recalling the conditional distribution of η_p with respect to σ^2 (3.13), the quantity could be written as:

$$\frac{1}{\sqrt{n_1}}\eta_p|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\sqrt{n_1}\xi_1 + \sqrt{n_1}\Phi^{-1}(p)\sigma, \sigma^2\right). \tag{3.22}$$

Since from the (3.17) it is known that the posterior distribution of σ^2 is GIG, the result of Theorem 2.1 can be used to obtain:

$$\frac{1}{\sqrt{n_1}}\eta_p|\mathbf{w} \sim SMNG(\bar{\lambda}, \sqrt{\delta^2 + n_1 v^2}, \gamma, \bar{\beta}, \sqrt{n_1}\xi_1). \tag{3.23}$$

To complete the proof, it is required to apply the result on the location-scale behaviour of the SMNG distribution studied in proposition 2.1.
(ii) Since $\eta_p|\mathbf{w}$ is SMNG distributed and $\theta_p = \exp\{\eta_p\}$, then by definition 2.2 it is log-SMNG distributed. ■

The deduced parameters of the θ_p posterior distribution are compliant with the meaning of the SMNG parameters: a higher posterior sample size n_1 implies lighter tails (smaller and negative λ and bigger γ) and a density that is gathered around the mode (smaller δ). On the other hand, the asymmetry parameter β is ruled by the studied quantile through the inverse of the standardized Gaussian cumulative distribution function, and the location parameter is equal to the conditioned posterior mean ξ_1.
As already shown in section 2.5, it is possible to deduce the moments of the log-SMNG distribution by starting from the moment generating function of the SMNG distribution. Since one of the aims of this work is to get a point estimate of the target quantity (Lehmann and Casella, 2006; Robert, 2007), the Bayes estimator associated to a given loss function, according to definition C.2, is investigated and evaluated. To have a brief introduction to Bayesian point estimation under loss functions see section C.
This kind of estimators represents a convenient way to synthesize the posterior distribution. In particular, the quadratic loss function (C.4) is considered in this work for its popularity

and the relative quadratic loss function (C.6) is included too because of its history in the context of Bayesian log-normal estimation. This aspect will be deepened in the following sections. The Bayes estimators associated to the considered loss functions respectively are the posterior mean and the ratio of posterior expectations reported in equation (C.7).

Proposition 3.3 (Bayes estimators of θ_p)**.** *Given that the posterior distribution for the target functional θ_p is (3.21), then:*

(i) the Bayes estimator of the log-normal p-th quantile under the quadratic loss function is:

$$\begin{aligned}\hat{\theta}_p^{QB} =& e^{\xi_1} \frac{\left(\frac{\sqrt{n_1}\gamma}{\sqrt{n_1\gamma^2-1}}\right)^{\bar{\lambda}}}{K_{\bar{\lambda}}(\sqrt{nv^2+\delta^2}\gamma)} \sum_{j=0}^{+\infty} \frac{\bar{\beta}^j}{j!} \left(\frac{\sqrt{nv^2+\delta^2}}{\sqrt{n_1}\sqrt{n_1\gamma^2-1}}\right)^{\frac{j}{2}} \times \\ & \times K_{\bar{\lambda}+\frac{j}{2}}\left(\frac{\sqrt{nv^2+\delta^2}\sqrt{n_1\gamma^2-1}}{\sqrt{n_1}}\right),\end{aligned} \tag{3.24}$$

that exists when $\gamma > \frac{1}{\sqrt{n}}$;

(ii) the Bayes estimator under relative quadratic loss is:

$$\begin{aligned}\hat{\theta}_p^{RQB} =& e^{\xi_1}\left(\frac{\sqrt{n_1\gamma^2-4}}{\sqrt{n_1\gamma^2-1}}\right)^{\bar{\lambda}} \times \\ & \times \frac{\sum_{j=0}^{+\infty}\frac{\bar{\beta}^j}{j!}\left(\frac{\sqrt{nv^2+\delta^2}}{\sqrt{n_1}\sqrt{n_1\gamma^2-1}}\right)^{\frac{j}{2}} K_{\bar{\lambda}+\frac{j}{2}}\left(\frac{\sqrt{nv^2+\delta^2}\sqrt{n_1\gamma^2-1}}{\sqrt{n_1}}\right)}{\sum_{j=0}^{+\infty}\frac{\bar{\beta}^j}{j!}\left(\frac{4\sqrt{nv^2+\delta^2}}{\sqrt{n_1}\sqrt{n_1\gamma^2-4}}\right)^{\frac{j}{2}} K_{\bar{\lambda}+\frac{j}{2}}\left(\frac{\sqrt{nv^2+\delta^2}\sqrt{n_1\gamma^2-4}}{\sqrt{n_1}}\right)};\end{aligned} \tag{3.25}$$

that exists when $\gamma > \frac{2}{\sqrt{n_1}}$.

Besides, to have a finite posterior variance it must hold that:

$$\gamma > \frac{2}{\sqrt{n_1}}. \tag{3.26}$$

Proof. (i) Since the Bayes estimator under quadratic loss is the posterior mean, considering that $\theta_p|\mathbf{w}$ follows a log-SMNG distribution, the result is a simple application of the proposition 2.3.
(ii) To get the Bayes estimator under relative quadratic loss it is sufficient to note that by applying proposition 2.1:

$$-\log(\theta_p)|\mathbf{w} \sim SMNG(\bar{\lambda}, \bar{\delta}, \bar{\gamma}, \bar{\beta}, -\xi_1), \tag{3.27}$$

$$-2\log(\theta_p)|\mathbf{w} \sim SMNG(\bar{\lambda}, 2\bar{\delta}, \bar{\gamma}/2, \bar{\beta}, -2\xi_1); \tag{3.28}$$

and then it is required to use proposition 2.3 to get an expression for (C.7).

■

Unfortunately, the obtained estimators are expressible only as an infinite sum of Bessel K functions. However, the convergence of the sum has been explored in chapter 2 and it can be easily implemented in any statistical software. In particular, in section 2.6 all the tools to perform the posterior analysis on a log-SMNG distribution are provided.
As expected, the existence of the Bayes estimators is subjected to a restriction on the parameter γ, the one which controls the right tail of the distribution. It is a less restrictive condition, if compared to the one deduced by Fabrizi and Trivisano (2012) for the log-normal mean. This is an expected result recalling the comparison between the SMNG distribution and the GH distribution (section 2.4). In fact, the functional involved in the quantile estimation is characterized by less variability and, consequently, lighter tails, because of the presence of σ instead of σ^2. Finally, the restriction on γ becomes negligible if n is increasing. Another appealing property of the existence condition is that it does not depend on the quantile estimated.
On the other hand, the existence condition points out the reason why the usual non-informative priors (like the Jeffrey's prior) do not produce a posterior distribution with finite moments. In fact, they might be deduced as particular cases of the GIG distribution with $\gamma \to 0$.
Furthermore, this result justifies the necessity of a GIG prior on the variance σ^2 instead of the most common inverse gamma distribution, that is the limiting case of a GIG distribution with $\gamma \to 0$. In fact, in the latter case, the posterior of θ_p would assume the exponential transformation of the random variable having density (2.33), whose mean is not finite (see section 2.3.3).

3.2.1 Minimum MSE conditional estimator

Following the idea of the paper by Zellner (1971), the Bayes estimator of θ_p conditioned with respect to σ^2 with minimum MSE might be found. In particular, the general class of estimators defined as: $\theta_p^* = k \cdot \exp\{\bar{w}\}$, where k is a constant, is considered. A vague improper prior is assumed for ξ:

$$p(\xi) \propto 1. \tag{3.29}$$

Even if the result cannot be applied in practice, it can be used as a benchmark to evaluate the performances of the proposed Bayes estimators.
The principal results about this particular estimator are contained in the following theorem.

Theorem 3.2. *Considering the estimators of the functional θ_p that consider σ^2 as known and are included in the class:*

$$\theta_p^* = k \cdot \exp\{\bar{w}\}; \tag{3.30}$$

then the one that minimizes the frequentist MSE is:

$$\hat{\theta}_p^* = \exp\left\{\bar{w} + \sigma\Phi^{-1}(p) - \frac{3\sigma^2}{2n}\right\}. \tag{3.31}$$

Furthermore, it coincides with the conditioned Bayes estimator that minimizes the relative quadratic loss function.

Proof. Plugging the generic form of the estimator (3.30) into the definition of MSE:

$$\begin{aligned}\mathbb{E}\left[\left(\theta_p^* - \theta_p\right)^2\right] &= k^2\mathbb{E}\left[\exp\left\{2\bar{w}\right\}\right] - 2k\mathbb{E}\left[\exp\left\{\bar{w}\right\}\right]\exp\left\{\xi + \sigma\Phi^{-1}(p)\right\} + \\ &\quad + \exp\left\{2\xi + 2\sigma\Phi^{-1}(p)\right\} \\ &= k^2\exp\left\{2\left(\xi + \frac{\sigma^2}{n}\right)\right\} - 2k\exp\left\{2\xi + \frac{\sigma^2}{2n} + \sigma\Phi^{-1}(p)\right\} + \\ &\quad + \exp\left\{2\xi + 2\sigma\Phi^{-1}(p)\right\}.\end{aligned} \tag{3.32}$$

It is immediate to see that this parabola is minimized when:

$$k = \exp\left\{\sigma\Phi^{-1}(p) - \frac{3\sigma^2}{2n}\right\}. \tag{3.33}$$

Then, if the relative quadratic loss function (C.6) is taken into account:

$$\begin{aligned}L &= \left[\frac{\theta_p - \theta_p^*}{\theta_p}\right]^2 \\ &= \left[1 - \theta_p^*\exp\left\{-\xi - \sigma\Phi^{-1}(p)\right\}\right]^2;\end{aligned} \tag{3.34}$$

the expectation taken with respect to $\xi|\sigma^2, \mathbf{w}$ is:

$$\begin{aligned}\mathbb{E}_\xi\left[L|\sigma^2, \mathbf{w}\right] =& 1 - 2\theta_p^*\mathbb{E}_\xi\left[e^{-\xi}|\sigma^2, \mathbf{w}\right]e^{-\sigma\Phi^{-1}(p)} + \\ &+ \theta_p^{*2}\mathbb{E}_\xi\left[e^{-2\xi}|\sigma^2, \mathbf{w}\right]e^{-2\sigma\Phi^{-1}(p)}.\end{aligned} \tag{3.35}$$

If the expectation of $e^{-\xi}$ and $e^{-2\xi}$ is derived recalling that in case of improper prior it holds:

$$\xi|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\bar{w}, \frac{\sigma^2}{n}\right), \tag{3.36}$$

then $\hat{\theta}_p^*$ is the value that minimizes the expression. ■

Therefore, an implicit result of this section is that, conditionally on σ^2, the posterior expectation of (3.14) does not minimize the frequentist MSE; since its minimum is reached by the Bayes estimator under relative quadratic loss. Furthermore, this result clarifies the inclusion of $\hat{\theta}_p^{RQB}$ in proposition 3.3.

3.3 Choice of hyperparameters

Once the prior distribution is fixed and all the posterior distributions are deduced, the following step is the hyperparameters selection. In fact, it is evident that posterior inference depends on the values assumed by $(\xi_0, n_0, \lambda, \delta, \gamma)$, particularly with small samples.
Two different strategies will be presented in order to fix the hyperparameters: the first one represents a weakly informative choice, whereas the second one is aimed at minimizing the frequentist MSE, following the idea of Rukhin (1986).

All the proposals reported in the following sections will be based on the fundamental idea that the hyperparameters should be specified in order to ensure the existence of the Bayes estimators and the posterior variance.

3.3.1 Weakly informative prior

As regards to the prior on the mean in the log-scale ξ, the usual flat improper prior could be obtained setting $n_0 \to 0$, without affecting the existence condition of the posterior moments. In this case, a prior that assumes independence between σ^2 and ξ is considered:

$$\begin{aligned} &p(\xi) \propto 1, \\ &\sigma^2 \sim GIG(\lambda, \delta, \gamma). \end{aligned} \tag{3.37}$$

Within this framework, some of the distributional results reported in section 3.2 are subjected to little changes. In particular, it is possible to note that:

$$\begin{aligned} &\xi_1 = \bar{w}, \\ &n_1 = n, \\ &\bar{\lambda} = \lambda - \frac{n-1}{2}. \end{aligned} \tag{3.38}$$

Therefore, as an example, the conditional posterior distribution of the mean in the log-scale ξ becomes:

$$\xi|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\bar{w}, \frac{\sigma^2}{n}\right). \tag{3.39}$$

In general, to deduce the quantities of interest under this prior setting, it is sufficient to plug the values reported in (3.38) into the results of section 3.2.
The particular form of the posterior mode of σ^2 could represent a reasonable starting point to specify the GIG hyperparameters $(\lambda, \delta, \gamma)$. In fact, considering the posterior distribution of σ^2 (3.17), recalling the (2.4), its mode is:

$$Mo\left(\sigma^2|\mathbf{w}\right) = \frac{(\bar{\lambda}-1) + \sqrt{(\bar{\lambda}-1)^2 + (nv^2+\delta^2)\,\gamma^2}}{\gamma^2}. \tag{3.40}$$

If the first order expansion of the square root around a value c is considered (known also as Babylonese method):

$$\sqrt{c^2+m} = |c| + \frac{m}{2|c|}, \tag{3.41}$$

then it is possible to approximate the mode:

$$\begin{aligned} Mo\left(\sigma^2|\mathbf{w}\right) &\simeq \frac{(\bar{\lambda}-1) + |\bar{\lambda}-1| + \frac{\left(nv^2+\delta^2\right)\gamma^2}{2|\bar{\lambda}-1|}}{\gamma^2} \\ &= \frac{\left(nv^2+\delta^2\right)}{-2\lambda+n+1}, \end{aligned} \tag{3.42}$$

if it is assumed that $\bar{\lambda} - 1 < 0$, i.e. $\lambda < \frac{n+1}{2}$; a condition that does not represent a binding restriction. From the last expression, it is possible to observe that if λ and δ are respectively fixed equal to $\frac{1}{2}$ and a small value ε, the posterior mode of σ^2 would approximately be equal to the sample variance in the log-scale. In that case, an inverse Gaussian distribution is in fact specified as prior.
Similar results might be deduced considering the approximation of the posterior expectation proposed in Fabrizi and Trivisano (2012):

$$\mathbb{E}\left[\sigma^2|\mathbf{w}\right] \simeq \frac{\bar{\lambda} + \sqrt{\bar{\lambda}^2 + (nv^2 + \delta^2)\,\gamma^2}}{\gamma^2}. \tag{3.43}$$

By applying the same concepts used to deduce the (3.42) the following quantity is deduced:

$$\mathbb{E}\left[\sigma^2|\mathbf{w}\right] \simeq \frac{\left(nv^2 + \delta^2\right)}{-2\lambda + n - 1}; \tag{3.44}$$

if $\lambda = 0$ and, again, $\delta = \varepsilon$ are fixed, then the posterior expectation of σ^2 would be approximately equal to the unbiased estimator of the variance of data in the log-scale.
Both the approximations introduced do not provide information about the value which γ should assume. Since it is the parameter that controls the tail behaviour, it can be reasonably fixed equal to the minimum quantity that allows the existence of the Bayes estimator and its variance. Therefore, recalling the (3.26), it is reasonable to assume that:

$$\gamma_0 = \frac{2}{\sqrt{n}} + \varepsilon_\gamma. \tag{3.45}$$

The choice of ε_γ could influence the inferential procedure, especially in case of small sample size. In fact, a very small value might not be appropriate since it can produce explosive values for the highest moment defined. The proposal that will be adopted from now on is to fix $\varepsilon_\gamma = \frac{1}{\sqrt{n}}$, which corresponds to the quantity required to reach the existence threshold for the moment with the nearest higher order.
Thanks to the GIG distribution flexibility, setting $\lambda = 0$ or $\lambda = 0.5$, a prior distribution with heavy tails is obtained, nonetheless a posterior with finite moments up to the second order is produced. In fact, by looking at figure 3.1, it appears that the right tail of the GIG distribution decay slower than the one of the inverse gamma near to 0, but then they do not assume the form of an asymptote like the inverse gamma tail, since they markedly continue the decrease. In this way the posterior moments finiteness is preserved.

3.3.2 Minimum frequentist MSE estimators

Bayes estimators can be evaluated also in terms of frequentist properties, in order to compare their performances to frequentist estimators (Carlin and Louis, 2008). Therefore, it is possible to propose procedures that allow to choose hyperparameters that minimize the frequentist MSE. The deduced prior settings can be particularly useful when a small sample is available and a point estimation is required.

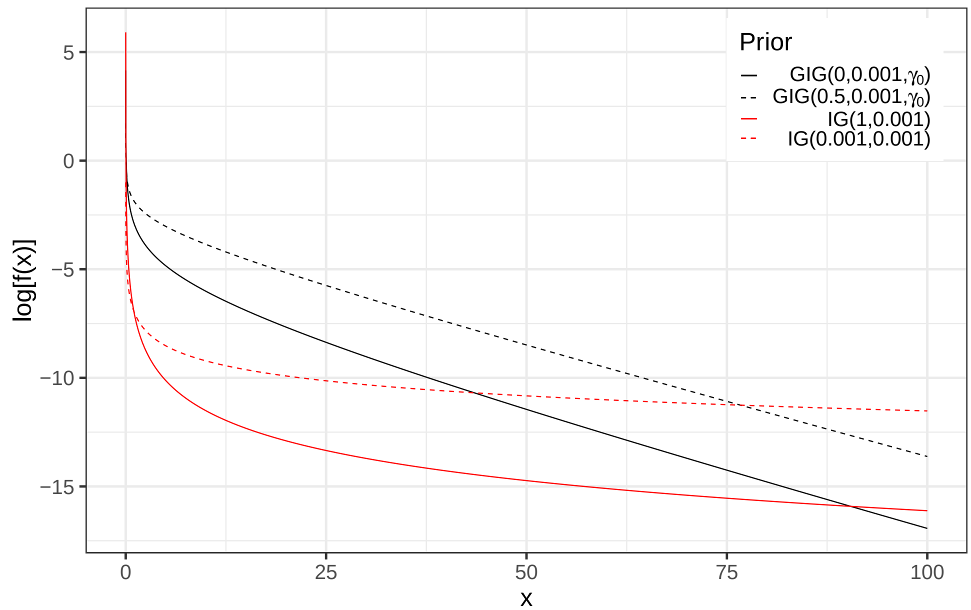

Figure 3.1: Log density of the weakly informative GIG distributions proposed and of the most common vague inverse gamma (IG) priors.

Since the proposed estimators have particularly complicated mathematical expressions, due to the presence of an infinite sum of Bessel K functions, closed form relations to find optimal parameters cannot be found, even if approximations are applied. Therefore, a numerical solution is firstly proposed. Afterwards, connections with results included in the paper by Fabrizi and Trivisano (2012) for the functional $\theta_{a,b}$ (1.4) are shown in order to have a strategy for the hyperparameters specification that does not involve a numerical software.

Numerical optimization

A five parameter optimization problem appears over-dimensioned for the inferential purpose, especially in a small sample framework.
Some suggestions to reduce the dimensionality of the minimizing function can be argued by specifying the MSE expression.
Recalling the (3.13), the MSE of the Bayes estimator under quadratic loss (3.24) can be decomposed in the following way:

$$\begin{aligned}\mathbb{E}\left[\left(\hat{\theta}_p^{QB}-\theta_p\right)^2\right] &= \mathbb{E}\left[e^{2(\psi\bar{w}+(1-\psi)\xi_0)}g(V^2)^2-2\theta_p e^{\psi\bar{w}+(1-\psi)\xi_0}g(V^2)+\theta_p^2\right]\\ &= \theta_p^2\left[e^{2(1-\psi)(\xi_0-\xi)+\frac{2\psi^2\sigma^2}{n}-2\phi\sigma}\mathbb{E}\left[\left(g(V^2)-e^{(1-\psi)(\xi_0-\xi)-\frac{3\psi^2\sigma^2}{2n}+\phi\sigma}\right)^2\right]+\right.\\ &\qquad \left.+1-e^{-\frac{\psi^2\sigma^2}{n}}\right],\end{aligned} \tag{3.46}$$

where $g(V^2)$ is a function of the sample variance V^2 and it is the only part of the expression that includes the hyperparameters of the GIG distribution.
A first way to simplify the research of the minimum is to specify n_0 in a weakly informative but proper way, e.g. to obtain $\psi = 0.98$. Through this decision, the estimator performance would be robust to misspecification of ξ_0, but the contribution of the normal prior instead of the flat prior might be useful in containing the estimator variance.
However, in this prior setting, the target functional to minimize with respect to the remaining parameters $(\lambda, \delta, \gamma)$ is approximately equal to:

$$\mathbb{E}\left[\left(g(V^2) - \exp\left\{\phi\sigma - \frac{3\sigma^2}{2n}\right\}\right)^2\right], \tag{3.47}$$

considering the (3.46) and letting $\psi \to 1$. This approximation coincides with the result that would be obtained in case of improper prior on ξ.
Therefore, the functional to minimize is an expectation taken with respect the random variable V^2, and from the log-normality assumption it holds that:

$$nV^2 \sim Ga\left(\frac{n-1}{2}, \frac{1}{2\sigma^2}\right). \tag{3.48}$$

By exploring the trend of function (3.47), the indication is that the searching problem is still over-dimensioned: it is not possible to find a unique minimum point unless two GIG parameters are kept constant.
Because of the necessity of putting constraints on the parameters range, numerical algorithms that allow to satisfy this requirement are considered. In particular, a bounds constrained quasi-Newton method that is implemented in the `R` package `optimx` (Nash et al., 2014) is employed.
The first parameter to fix is λ: it is a shape parameter and it appears in the order of the Bessel's K functions and an eventual numerical optimization algorithm for it might be too unstable. Following the idea of the previous section, the shape parameter λ is fixed equal to 0. By observing the trends of the (3.47) in this case, it appears that a finite minimum cannot still be found optimizing with respect two parameters. Besides, a distinction in the procedure for quantiles above the median and below the median (case that will be dealt with later) seems to be necessary. Moreover, the different behaviour of the functional over the quantiles can provide some suggestions in order to fix another parameter and to have the possibility of finding the third value through minimization (figure 3.2).
To capture the shape of the function the following strategies could be adopted:

- $p < 0.5$: fix γ equal to the minimum value that allows the existence of second posterior moment according to the rule in (3.45). Then minimize with respect to the parameter δ;

- $p > 0.5$: fix δ and minimize with respect to γ. Recalling the (3.44), it is possible to specify an informative value of δ, considering it as a contribution in terms of variance

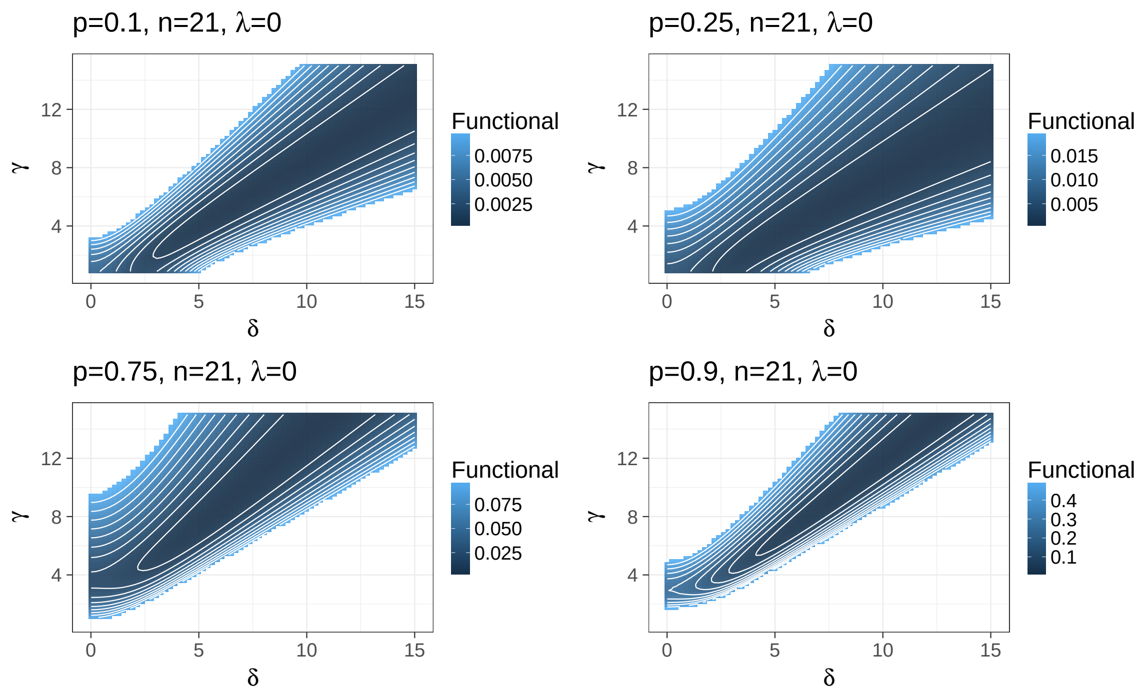

Figure 3.2: Behaviour of the functional in (3.47) for $\hat{\theta}_p^{QB}$ with respect to the parameters γ and δ, keeping constant $\lambda = 0$ and considering $p = (0.10, 0.25, 0.75, 0.90)$. The case $n = 21$ and $\sigma^2 = 1$ is considered.

of a prior sample. A general proposal could be $\delta = 1$: in most applied problems, values of the variance in the log-scale σ^2 are seldom greater than 2, so 1 can be read as a reasonable guess for the size of an hypothetical deviation from the mean when $n = 1$. Of course, if the scale of the problem is totally different, the user can specify alternative values for δ.

Heuristically, searching for optimal γ for quantiles above the median, and optimal δ for those below, is in line with the specialization of these parameters in the GIG distribution: γ rules the right tail of the distribution that is not relevant when $p < 0.5$, while δ is more involved with the general spread of the distribution and is therefore more relevant to the shape of the lower tail.
In the practical context, it must be considered that σ^2 is unknown and it appears in the functional (3.47). This issue might be overcome by plugging into the expression the sample variance v^2 or a guess s_0^2 if it is available. The latter procedure is advisable since it allows to remove from the MSE the part of variability caused by the use of v^2. Moreover, the procedure could be considered more rigorous from a Bayesian viewpoint, since data would not been used twice in the inferential procedure. However, it is not always possible to have a safe value, even if the procedure is quite robust to the misspecification of s_0^2, as will be in the simulation study presented in the following chapter.

A considerable particular case is the median and its neighbourhood. In these cases, the Bayes estimator under relative quadratic loss seems to perform better than the posterior mean for several reasons. Considering the figure 3.3, it could be deduced that the MSE is minimized when $\lambda \to -\infty$ is chosen. Using the limiting form (A.8) for the Bessel K function, it is possible to prove that the Bayes estimator under quadratic loss equates the naive estimator $\exp\{\bar{w}\}$ with that degenerate prior. As a consequence, the so called naive estimator for the median represents the best case for the Bayes estimator under quadratic loss.
On the contrary, the Bayes estimator under relative quadratic loss $\hat{\theta}_p^{RQB}$ presents a non-monotone decreasing trend of the MSE only around the median. For this reason it appears to be the preferable estimator in this situation, but the characteristics of $\hat{\theta}_p^{RQB}$ deteriorate in a fast way departing from the median. This particular behaviour is related to the absence of σ in $\theta_{0.5}$: it is a case similar to the conditional estimator of θ_p dealt with in section 3.2.1. In this framework, the Bayes estimator under relative quadric loss resulted the minimum MSE choice. Moreover, because of the absence of σ in the estimand, the hyperparameters specification does not influence the estimation step. Therefore the weakly informative prior could be used in that case.

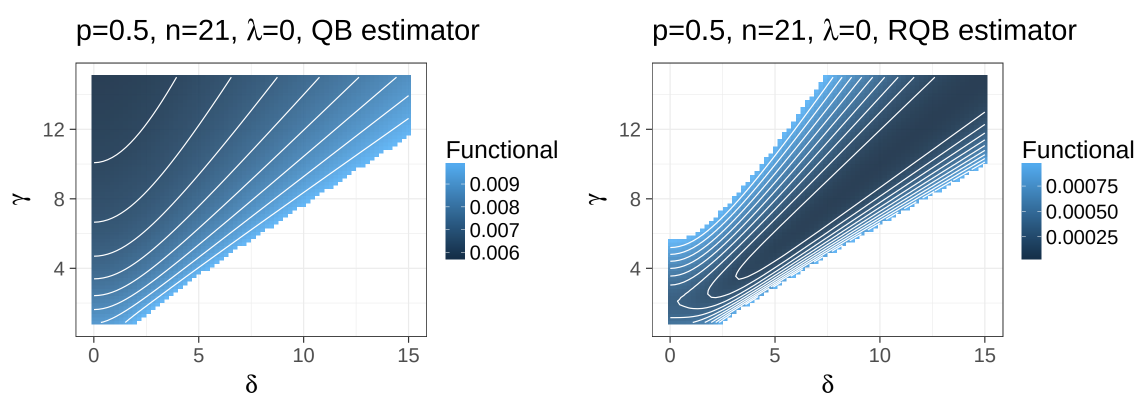

Figure 3.3: Behaviour of the functional in (3.47) for $\hat{\theta}_{0.5}^{QB}$ and $\hat{\theta}_{0.5}^{RQB}$ with respect to the parameters γ and δ, keeping constant $\lambda = 0$. The case $n = 21$ and $\sigma^2 = 1$ is considered.

Connection with the Bayes estimator of $\theta_{a,b}$

In the paper by Fabrizi and Trivisano (2012) the problem of the Bayesian inference of the functional $\theta_{a,b} = \exp\{a\xi + b\sigma^2\}$, assuming a log-normal distribution for data, was studied. In that context, the prior (3.37) was considered, and the found posterior distribution of $\theta_{a,b}$ was a GH. Furthermore, a strategy to find a minimum MSE Bayes estimator was proposed. Through an analytic approximation of the target functional to minimize, the optimal value of λ, for a fixed δ, was found to be:

$$\lambda_{opt} = \frac{n-3}{2} - \frac{(n-1)(a^2+2nb)}{4nc} - \frac{(a^2+2nb)\delta}{4nc\sigma^2}, \tag{3.49}$$

where $c = \left(b - \frac{3a^2}{2n}\right)$. Moreover, the following condition must hold: $b \notin \left(-\frac{a^2}{2n}, \frac{3a^2}{2n}\right)$. It is useful to set $\delta \to 0$ in order to remove the effect of σ^2, avoiding the substitution with the sample quantity v^2 or a guess s_0^2. The value of γ was selected in order to be sure of the second posterior moment existence. According to the theory of the GH distribution it is:

$$\gamma_0 = \max\left\{0, 4\left(\frac{a^2}{n} + b\right)\right\} + \varepsilon. \tag{3.50}$$

Note that if $a = 1$ and $b = 0$, the median of the log-normal distribution is considered, then: $\theta_{1,0} = \theta_{0.5}$. This case violates the condition on b required for the validity of relation (3.49). This is in agreement with the MSE trend showed in figure 3.3, where a finite minimum cannot be found. As a consequence, the indications included in this section cannot be applied in the median estimation context.

In order to apply the relation (3.49) in the log-normal quantile estimation problem, the proposal is based on the idea that each functional specified by a couple (a, b) corresponds to a quantile θ_p too. Utilizing the similarity among the quantiles that are above or below the median, the relation (3.49) could be used to obtain a unique set of optimal values for all the quantiles that belongs to the same *half* cumulative distribution (i.e. $p < 0.5$ or $p > 0.5$).

Even if the (3.49) is a function of n, the sample size does not significantly change the final result. An empirical general choice for λ could be -2 for the quantiles above the median, and 0.5 for the quantiles below the median, always keeping $\delta = \varepsilon$ and gamma equal to (3.45). For example, considering that in this case $b = \frac{\Phi^{-1}(p)}{\sigma}$, they correspond, with the intermediate value $\sigma^2 = 1$ to the optimal value for $p = 0.85$ and $p = 0.30$, respectively.

This method produces surprisingly good performances in terms of frequentist MSE, as will be stressed in the following chapter about the simulation study, and it possesses an appealing property if the statistical analysis goal is the joint estimation of different quantiles: since a unique prior is specified, the user is sure that the estimation procedure maintains the logical order of the quantiles and counter-intuitive results are avoided.

Prior proposals: brief outline and software implementation

In the previous sections different prior specifications were listed. Using figure 3.4 it is possible to sum up the proposals and fix some notation. Starting from the weakly informative prior, the triplet $(\lambda = 0, \delta = \varepsilon, \gamma = \gamma_{min})$ produces the posterior mean of the p-th log-normal quantile $\hat{\theta}_p^{QBw}$, but is also used in the median estimations issue with the Bayes estimator under relative quadratic loss $\hat{\theta}_p^{RQBw}$. The value $\lambda = 0.5$ might be used too.

Then, a frequentist optimal (i.e. minimum MSE) Bayes estimator is considered, and an optimization algorithm is used for the hyperparameters choice. In this case the Bayes estimator under quadratic loss associated to the prior setting is $\hat{\theta}_p^{QBn}$ and in figure 3.4 priors examples are reported considering: $p = 0.10$ and $p = 0.25$ below the median (optimization with respect to δ) and $p = 0.75$ and $p = 0.90$ above the median (optimization with respect to γ).

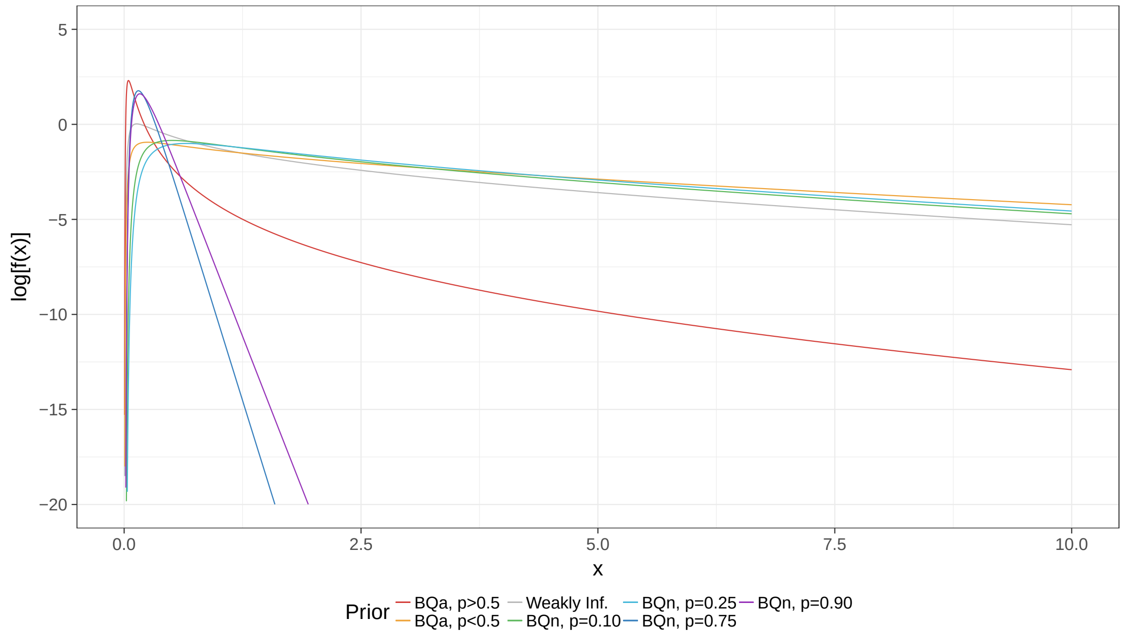

Figure 3.4: Comparison of log-densities of different GIG priors. A random sample of size $n = 21$ generated from a log-normal having $\sigma^2 = 0.5$ is used.

Finally, an *approximately optimal* Bayes estimator is proposed using the similarities between θ_p and $\theta_{a,b}$. Two generic triplets are proposed for the lower and upper parts of the distribution. In this case the Bayes estimator is specified with $\hat{\theta}_p^{QBa}$.
In the package `BayesLN` the function `LN_Quant` is provided in order to allow the user to carry out a Bayesian inferential procedure on the log-normal quantiles. To simplify the usage of the function, two prior specification settings are only proposed: the use of $\lambda = 0$ for the weakly informative prior, and the optimal prior obtained thorough numerical minimization of the MSE. Concerning the prior on ξ, a flat improper prior is considered.

3.4 Extension to the regression case

In many applications, it is useful to estimate a log-normal linear regression model for the p-th quantile of the dependent variable. This inferential problem is usually faced with non-parametric techniques both in frequentist and Bayesian worlds (Gilchrist, 2000; Yu and Moyeed, 2001; Koenker, 2005). However, when positively skewed data with scarce sample sizes is analysed (Vogel et al., 2011; Machado et al., 2015), the log-normality assumption is helpful in gaining efficiency in the estimation, and a better procedure that the usual exponentiation of estimates obtained in the log-scale might be of interest.
On the other hand, if strictly positive data are analysed, the so called Box-Cox quantile regression is employed to keep the correct range for the predicted values (Yu et al., 2003;

Fitzenberger et al., 2009). The procedure consists in the joint estimation of the transformation parameters and the classical model-free quantile regression (Koenker and Bassett, 1978). Then, the results are back-transformed to the original data scale.
In this context, a random sample is observed:

$$(y_i, \mathbf{x}_i),\ i = 1 \dots n;$$

where $\mathbf{x}_i$ is a vector containing the values of the p covariates that are related to the i-th unit. Besides, the vector of the logarithmic transformation of the response variable is $\mathbf{w} = \log(\mathbf{y})$. Finally, the following distributional assumption is fixed:

$$y_i|\mathbf{x}_i, \boldsymbol{\beta}, \sigma^2 \sim \log\mathcal{N}\left(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma^2\right),\ i = 1, \dots n. \tag{3.51}$$

In this case, the inferential question that will be answered is the estimation of the p-th quantile given a point $\tilde{\mathbf{x}} \in \mathbb{R}^q$ of the covariate space:

$$\theta_p(\tilde{\mathbf{x}}) = \mathbb{Q}_p\left[\tilde{y}|\tilde{\mathbf{x}}\right] = \exp\left\{\tilde{\mathbf{x}}^T\boldsymbol{\beta} + \Phi^{-1}(p)\sigma\right\}. \tag{3.52}$$

Consistently with the previous sections, it is possible to express the logarithmic transformation of the functional $\theta_p(\tilde{\mathbf{x}})$ as $\eta_p(\tilde{\mathbf{x}})$.
To make inference on the quantities of interest, the standard NGIG conjugate prior is assumed:

$$\boldsymbol{\beta}|\sigma^2 \sim MV\mathcal{N}_q\left(\boldsymbol{\beta}_0, \sigma^2\mathbf{V}_0\right), \tag{3.53}$$
$$\sigma^2 \sim GIG(\lambda, \delta, \gamma); \tag{3.54}$$

where $\boldsymbol{\beta}_0 \in \mathbb{R}^q$, $\mathbf{V}_0 \in \mathbb{R}^{q\times q}$ and positive definite are chosen constants.
The appealing characteristic of this formulation is the generality, since most of the widespread normal regression priors (Zellner's g, semi-conjugate and flat) can be seen as particular cases.
The following proposition includes a series of results conditioned on σ^2 that derive from the traditional Bayesian analysis of the normal linear model.

Proposition 3.4. *If the conjugate prior for* $\boldsymbol{\beta}$ (3.53) *and the log-normal regression model* (3.51) *are assumed, then the following distributional results, conditionally on* σ^2*, can be deduced:*

(i)

$$\boldsymbol{\beta}|\sigma^2, \mathbf{w} \sim MV\mathcal{N}_q\left(\boldsymbol{\beta}_*, \sigma^2\mathbf{V}_*\right), \tag{3.55}$$

where $\boldsymbol{\beta}_* = \mathbf{V}_*\left(\mathbf{X}^T\mathbf{w} + \mathbf{V}_0^{-1}\boldsymbol{\beta}_0\right)$ *and* $\mathbf{V}_* = \left(\mathbf{X}^T\mathbf{X} + \mathbf{V}_0^{-1}\right)^{-1}$*;*

(ii)

$$\tilde{\mathbf{x}}^T\boldsymbol{\beta}|\sigma^2, \mathbf{w} \sim \mathcal{N}\bigg(\tilde{\mathbf{x}}^T\boldsymbol{\beta}_*, \sigma^2\tilde{h}_*\bigg), \tag{3.56}$$

where $\tilde{h}_* = \tilde{\mathbf{x}}^T\mathbf{V}_*\tilde{\mathbf{x}}$*;*

(iii)

$$\eta_p(\tilde{\mathbf{x}})|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\tilde{\mathbf{x}}^T\boldsymbol{\beta}_* + \Phi^{-1}(p)\sigma, \sigma^2\tilde{h}_*\right). \tag{3.57}$$

Proof. The three results are deduced from the joint posterior distribution:

$$\begin{aligned} p(\boldsymbol{\beta}, \sigma^2|y) &\propto p(\mathbf{w}|\boldsymbol{\beta}, \sigma^2)p(\boldsymbol{\beta}|\sigma^2)p(\sigma^2) \\ &\propto (\sigma^2)^{\lambda - \frac{n+q}{2} - 1} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}_*)^T \mathbf{V}_*^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_*)\right\} \times \\ &\times \exp\left\{-\frac{1}{2}\left[\frac{RSS + \delta^2 + \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)^T \mathbf{X}^T\mathbf{X}\mathbf{V}_*\mathbf{V}_0^{-1}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sigma^2} + \gamma^2\sigma^2\right]\right\}, \end{aligned} \tag{3.58}$$

that is obtained by applying the algebra of the quadratic forms and defining: $RSS = \left(\mathbf{w} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)^T\left(\mathbf{w} - \mathbf{x}\hat{\boldsymbol{\beta}}\right)$, with $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{w}$. ■

Besides, in the last proposition, a key quantity for the log-normal linear regression has been introduced, that is the leverage $\tilde{h}_*$. In fact, from the (3.57), and the direct consequence that $\theta_p(\tilde{\mathbf{x}})$ is log-normally distributed, it is clear that the log-normal regression is an intrinsically heteroskedastic model in data scale.
These results are preparatory to define the marginal distributions of the parameters and the functionals that need to be estimated.

Theorem 3.3. *If the conjugate prior* (3.53), (3.54) *and the log-normal linear regression model* (3.51) *are assumed, then the following distributional results can be deduced:*

(i)

$$\sigma^2|\mathbf{w} \sim GIG\left(\bar{\lambda}, \bar{\delta}, \gamma\right), \tag{3.59}$$

where $\bar{\lambda} = \lambda - \frac{n-1}{2}$ *and* $\bar{\delta} = \sqrt{RSS + \delta^2 + \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)^T \mathbf{X}^T\mathbf{X}\mathbf{V}_*\mathbf{V}_0^{-1}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}$.

(ii)

$$\boldsymbol{\beta}|\mathbf{w} \sim MVGH_q\left(\bar{\lambda}, \gamma, \boldsymbol{\beta}_*, \mathbf{V}_*, \bar{\delta}, \mathbf{0}\right) \tag{3.60}$$

(iii)

$$\theta_p(\tilde{\mathbf{x}})|\mathbf{w} \sim \log SMNG\left(\bar{\lambda}, \bar{\delta}\sqrt{\tilde{h}_*}, \frac{\gamma}{\sqrt{\tilde{h}_*}}, \bar{\beta}, \tilde{\mathbf{x}}^T\boldsymbol{\beta}_*\right), \tag{3.61}$$

where $\bar{\beta} = \frac{\Phi^{-1}(p)}{\sqrt{\tilde{h}_*}}$.

Proof. *(i)* The marginal posterior distribution of σ^2 is obtained integrating out $\boldsymbol{\beta}$ from the joint posterior (3.58):

$$\begin{aligned} p(\sigma^2|\mathbf{w}) \propto &(\sigma^2)^{\lambda - \frac{n}{2} - 1} \times \\ &\times \exp\left\{-\frac{1}{2}\left[\frac{RSS + \delta^2 + \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)^T \mathbf{X}^T\mathbf{X}\mathbf{V}_*\mathbf{V}_0^{-1}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)}{\sigma^2} + \gamma^2\sigma^2\right]\right\}. \end{aligned} \tag{3.62}$$

and the kernel of a GIG distribution can be recognized. *(ii)* On the other hand, the marginal posterior distribution for $\boldsymbol{\beta}$ might be obtained using the result by Barndorff-Nielsen (1977): in fact it is a normal mean-variance mixture with the (3.55) as conditional multivariate Gaussian and the GIG of point *(i)* as mixing distribution, then a multivariate GH distribution (see section B.1.1) is obtained.
(iii) Finally, in order to have the posterior distribution of $\eta_p(\tilde{\mathbf{x}})$, it is required to consider the linear transformation of the (3.57):

$$\sqrt{\tilde{h}_*} \cdot \eta_p(\tilde{\mathbf{x}})|\sigma^2, \mathbf{w} \sim \mathcal{N}\left(\frac{\tilde{\mathbf{x}}^T \boldsymbol{\beta}_*}{\sqrt{\tilde{h}_*}} + \frac{\Phi^{-1}(p)}{\sqrt{\tilde{h}_*}}\sigma, \sigma^2\right), \tag{3.63}$$

and the (3.59), obtaining a SMNG distribution, according to theorem 2.1:

$$\sqrt{\tilde{h}_*} \cdot \eta_p(\tilde{\mathbf{x}})|\mathbf{w} \sim SMNG\left(\bar{\lambda}, \bar{\delta}, \gamma, \bar{\beta}, \frac{\tilde{\mathbf{x}}^T \boldsymbol{\beta}_*}{\sqrt{\tilde{h}_*}}\right); \tag{3.64}$$

and, by applying a linear transformation (proposition 2.1):

$$\eta_p(\tilde{\mathbf{x}})|\mathbf{w} \sim SMNG\left(\bar{\lambda}, \bar{\delta}\sqrt{\tilde{h}_*}, \frac{\gamma}{\sqrt{\tilde{h}_*}}, \bar{\beta}, \tilde{\mathbf{x}}^T \boldsymbol{\beta}_*\right). \tag{3.65}$$

■

Then, the definition of the Bayes estimators under quadratic loss and under relative quadratic loss is a consequence of these results, obtaining the parallel of proposition 3.3.

Proposition 3.5. *If the conjugate prior* (3.53),(3.54) *and the log-normal regression model* (3.51) *are assumed, then the Bayes estimator under relative quadratic loss is the posterior mean:*

$$\hat{\theta}_p^{QB}(\tilde{\mathbf{x}}) = e^{\tilde{\mathbf{x}}^T \hat{\boldsymbol{\beta}}} \frac{\left(\frac{\gamma}{\sqrt{\gamma^2 - \tilde{h}_*}}\right)^{\bar{\lambda}}}{K_{\bar{\lambda}}(\bar{\delta}\gamma)} \sum_{j=0}^{+\infty} \frac{\bar{\beta}^j}{j!} \left(\frac{\bar{\delta}\tilde{h}_*}{\sqrt{\gamma^2 - \tilde{h}_*}}\right)^{\frac{j}{2}} K_{\bar{\lambda}+\frac{j}{2}}\left(\bar{\delta}\sqrt{\gamma^2 - \tilde{h}_*}\right); \tag{3.66}$$

whereas the Bayes estimator under relative quadratic loss is:

$$\begin{aligned} \hat{\theta}_p^{RQB}(\tilde{\mathbf{x}}) =& e^{\tilde{\mathbf{x}}^T \hat{\boldsymbol{\beta}}} \left(\frac{\sqrt{\gamma^2 - 4\tilde{h}_*}}{\sqrt{\gamma^2 - \tilde{h}_*}}\right)^{\bar{\lambda}} \times \\ & \times \frac{\sum_{j=0}^{+\infty} \frac{\bar{\beta}^j}{j!} \left(\frac{\bar{\delta}\tilde{h}_*}{\sqrt{\gamma^2 - \tilde{h}_*}}\right)^{\frac{j}{2}} K_{\bar{\lambda}+\frac{j}{2}}\left(\bar{\delta}\sqrt{\gamma^2 - \tilde{h}_*}\right)}{\sum_{j=0}^{+\infty} \frac{\bar{\beta}^j}{j!} \left(\frac{4\bar{\delta}\tilde{h}_*}{\sqrt{\gamma^2 - 4\tilde{h}_*}}\right)^{\frac{j}{2}} K_{\bar{\lambda}+\frac{j}{2}}\left(\bar{\delta}\sqrt{\gamma^2 - 4\tilde{h}_*}\right)}; \end{aligned} \tag{3.67}$$

and in order to ensure the existence of the posterior variance the following condition is required:

$$\gamma > 2\sqrt{\tilde{h}_*}. \tag{3.68}$$

Proof. See proposition 3.3. ■

The results obtained in this section include the unconditional quantile estimation as particular case, since in that case $\tilde{h}_* = \frac{1}{n}$.

Minimum MSE conditional estimator

A minimum MSE estimator conditioned on σ^2, similar to the one deduced in section 3.2.1, could be found also for the quantile regression case. It is useful as a benchmark and it might be found for the general class of estimators defined as: $\theta_p^*(\tilde{\mathbf{x}}) = k \cdot \exp\{\tilde{\mathbf{x}}^T \hat{\boldsymbol{\beta}}\}$, where k is a constant and an improper flat prior is assumed for $\boldsymbol{\beta}$:

$$p(\boldsymbol{\beta}) \propto 1. \tag{3.69}$$

The result parallel to theorem 3.2 is included in the following proposition.

Proposition 3.6. *Considering the estimators of the functional* (3.52) *that assumes σ^2 known and are included in the class:*

$$\theta_p^*(\tilde{\mathbf{x}}) = k \cdot \exp\{\tilde{\mathbf{x}}^T \hat{\boldsymbol{\beta}}\}, \tag{3.70}$$

then, the one that minimize the frequentist MSE is:

$$\hat{\theta}_p^*(\tilde{\mathbf{x}}) = \exp\left\{\bar{w} + \sigma\Phi^{-1}(p) - \frac{3\tilde{h}\sigma^2}{2}\right\}, \tag{3.71}$$

where $\tilde{h}$ is the leverage in the vague improper prior setting:

$$\tilde{h} = \tilde{\mathbf{x}}\left(\mathrm{X}^T\mathrm{X}\right)^{-1}\tilde{\mathbf{x}}. \tag{3.72}$$

Furthermore, it coincides with the conditioned Bayes estimator that minimizes the relative quadratic loss function.

Proof. See theorem 3.2, recalling the (3.56). ■

3.4.1 Choice of the hyperparameters

Like in the unconditional log-normal quantiles estimation, inference depends on the hyperparameters specification, particularly with a small sample sizes.

In order to simplify the prior choice issue, the popular improper diffuse prior is assumed for the vector of regression coefficient $\boldsymbol{\beta}$. This decision leads to the following parameters for the posterior distribution of σ^2 (3.59):

$$\begin{aligned}
\bar{\delta} &= \sqrt{RSS + \delta^2}, \\
\bar{\lambda} &= \lambda - \frac{n-q}{2}, \\
\mathbf{V}_* &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}, \\
\boldsymbol{\beta}_* &= \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)\mathbf{X}^T\mathbf{w};
\end{aligned} \tag{3.73}$$

recalling that the new definition of the leverage was already presented in the (3.72).
The remaining parameters to specify are $(\lambda, \delta, \gamma)$, which control the GIG prior on σ^2. Different strategies can be proposed and following the same scheme of section 3.3 a weakly informative setting is firstly considered and then two procedures aimed at proposing an efficient estimator are described. The notation and the computations developed so far consider a single point estimation problem, characterized by a covariate vector $\tilde{\mathbf{x}}$. However, if the interest is in the joint estimation of $\tilde{n}$ quantiles, it is possible to store the covariate patterns in the rows of the matrix $\tilde{\mathbf{X}} \in \mathbb{R}^{q\times\tilde{n}}$.

Weakly informative prior

Reflecting the considerations reported in section 3.3.1 and in the work by Fabrizi and Trivisano (2016), a non-informative prior could be specified fixing $\lambda = 0$, $\delta = \epsilon$ and the smaller γ that allows the existence of the posterior variance:

$$\gamma_0 = 3\sqrt{\tilde{h}_M}, \tag{3.74}$$

where $\tilde{h}_M = \max_{i=1,\ldots,\tilde{n}}\{\tilde{h}_{ii}\}$, i.e. the maximum leverage of the observations to estimate, obtained substituting the i-th row of the matrix $\tilde{\mathbf{X}}$ into the (3.72).

Numerical Optimization

If the goal is to obtain an estimator with good frequentist properties, the MSE should be considered. Considering the estimation of a single unit and specifying a diffuse improper prior on the regression coefficients, the MSE expression of the Bayes estimator of $\theta_p(\tilde{\mathbf{x}})$ could be decomposed as:

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{\theta}_p(\tilde{\mathbf{x}}) - \theta_p(\tilde{\mathbf{x}})\right)^2\right] = \theta_p(\tilde{\mathbf{x}})^2 \exp\left\{-2\Phi^{-1}(p)\sigma + 2\tilde{h}\sigma^2\right\} \\
\left[\mathbb{E}\left(g(RSS) - \exp\left\{\Phi^{-1}(p)\sigma - \frac{3}{2}\tilde{h}\sigma^2\right\}\right)^2 + \right. \\
\left. + 1 - \exp\left\{-\tilde{h}\sigma^2\right\}\right].
\end{aligned} \tag{3.75}$$

In this expression, a quantity that is similar to the (3.47) is individuated and it is the unique fraction of the MSE that depends on the parameters to chose. Therefore, a numerical optimization could be implemented in order to get hyperparameters that minimizes the frequentist MSE. It is important to note that both the target functional to minimize and the estimator include the leverage point $\tilde{h}$. The use of a numerical algorithm when the observation specific leverage assumes particularly high values is deprecated since the method produces unstable evaluations of the optimal parameters. As a consequence, the proposal is the substitution of $\tilde{h}$ into the (3.75) and the (3.66) with the average leverage of the observed sample:

$$\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n} = \frac{q}{n}. \tag{3.76}$$

This solution solves also the issues that arise if a multiple units estimation problem is faced. By observing the surfaces representing the behaviour with respect to the hyperparameters of the target functional to minimize (similar to those in figure 3.2 and not reported here), it is possible to conclude that the same rules of section 3.3.2 could be adapted for this issue too: for the quantiles above the median, an optimal γ if searched keeping fixed $\lambda = 0$ and $\delta = 1$, whereas for the quantiles below the median the function is optimized with respect to δ with $\lambda = 0$ and $\gamma = 3\sqrt{\tilde{h}_M}$, considering a multiple estimation problem. Finally, also in this case, the posterior mean does not appear to be a good choice in minimizing the MSE in the median estimation framework, and the Bayes estimator under relative quadratic loss should be considered for this task.

Connection with the Bayes estimator of $\theta_{a,b}(\tilde{\mathbf{x}})$

Finally, another proposal for the hyperparameters selection is based on the analytical approximation found by Fabrizi and Trivisano (2016) for the conditioned log-normal mean estimation, with the following target functional: $\theta_{1,\frac{1}{2}}(\tilde{\mathbf{x}}) = \exp\left\{\tilde{\mathbf{x}}\boldsymbol{\beta} + \frac{1}{2}\sigma^2\right\}$. In analogy with the procedure in section 3.3.2, it is possible to consider the log-normal mean as a quantile above the median, and therefore, the idea is to assume that an optimal parameter for its estimation is a value that produces good estimates also for the right tail quantiles. The same reasoning might be extended to quantiles in the left tail, considering the functional $\theta_{1,-\frac{1}{2}}(\tilde{\mathbf{x}}) = \exp\left\{\tilde{\mathbf{x}}\boldsymbol{\beta} - \frac{1}{2}\sigma^2\right\}$.

This approach ensures that the estimated conditional quantiles for the same tail do not cross. This represents a possible drawback of many quantile regression techniques, and a lot of work has been done in order to propose methods that are able to correct this undesirable property (Bondell et al., 2010; Chernozhukov et al., 2010). In this case, it would be a direct consequence of the parametric assumption together with the unique prior setting.

The procedure implemented was based on a small argument approximation of the MSE that leads to an optimal relation between the parameters (λ, δ). With the choice of $\delta = \varepsilon$, the value of λ does not depend on the estimated variance v^2, and therefore it is a suitable choice. The third parameter γ disappear in the approximation and could be fixed to the minimum value that ensures the existence of the posterior variance.

A remaining discussion point is about the presence of the leverage into the expression that determines λ. Coherently with the considerations about the numerical optimization prior, the average leverage $\bar{h}$ is a suitable choice. The relation used to find the optimal λ works if the condition $n > 3q$ is satisfied, that is usually verified also in the smallest samples.
To conclude, with the setting $\delta = \varepsilon$, the generic prior has λ equal to:

$$\lambda_{opt} = \begin{cases} \frac{n-q-2}{2} - \frac{(n-q)(q+n)}{2(n-3q)}, & p < 0.5 \\ \frac{n-q-2}{2} - \frac{(n-q)(q-n)}{2(-n-3q))}, & p > 0.5 \end{cases}, \tag{3.77}$$

and $\gamma_0 = 3\sqrt{\tilde{h}_M}$.

Software implementation

The function `LN_QuantReg`, which is included in the package `BayesLN`, allows to estimate a conditioned quantile under the assumption of log-normality. It supports the joint estimation of multiple points and it automatically uses the maximum leverage $\tilde{h}_M$ to set the existence condition affecting the parameter γ. As in the unconditional quantile estimation case, the weakly informative prior setting is implemented with $\lambda = 0$. On the other hand, the optimal prior setting, in which numerical procedures are utilized, is implemented using the average leverage $\bar{h}$ in the optimization step and in the estimator expression.

Chapter 4

Quantile estimation: simulations and examples

To assess the frequentist properties of the methods developed in chapter 3, a Monte Carlo simulation study is performed. The different strategies that are proposed to estimate the log-normal quantiles within the Bayesian paradigm are compared to Longford's minimum MSE estimator $\hat{Q}_p$ (section 3.1.3), the naive estimator $\hat{\theta}_p$ (section 3.1.2) and the non-parametric estimator $\hat{Q}_p^7$ (section 3.1.1). Several combinations of the log-scale variance σ^2, sample size n and quantile level p are investigated in the study. The location parameter ξ is set equal to 0 without loss of generality. In particular, the study will focus on small sample sizes, that is the condition in which the proposed Bayes estimators and credible intervals bring significant improvements to the currently used methods.
The evaluation of the estimators performance is carried out in terms of relative root-MSE, that is defined as follows for the generic estimator $\tilde{\theta}_p$:

$$RRMSE\left[\tilde{\theta}_p\right] = \frac{\sqrt{\mathbb{E}\left[\left(\tilde{\theta}_p - \theta_p\right)^2\right]}}{\theta_p}, \tag{4.1}$$

where the value of θ_p is obtained plugging the true quantity into the functional (1.5); and relative bias:

$$RB\left[\tilde{\theta}_p\right] = \frac{\mathbb{E}\left[\tilde{\theta}_p - \theta_p\right]}{\theta_p}. \tag{4.2}$$

Since θ_p is always positive, there are no issues in the definition of the relative bias.
In some situations, mainly for graphical purposes, the minimum MSE Bayes estimator $\hat{\theta}_p^*$, that assumes σ^2 as known (described in section 3.2.1), is used as a benchmark for RRMSE. In this cases a Comparative RMSE is used:

$$CRMSE\left[\tilde{\theta}_p\right] = \frac{RMSE\left[\tilde{\theta}_p\right]}{RMSE\left[\tilde{\theta}_p^*\right]}. \tag{4.3}$$

On the other hand, when the focus is on the interval estimation, the Bayesian credible intervals are compared in terms of frequentist coverage and average width to the ones derived from the Gaussian theory (section 3.1.2).
A preliminary note about the simulations involving the Bayes estimator with optimal hyperparameters which are selected through the numerical optimization (see section 3.3.2), is necessary. In fact, the algorithm is computationally demanding and the whole procedure cannot be repeated for each sample to have a reasonable duration of the simulations. The following procedure is adopted: since the target functional depends on v^2, a grid of 8 equispaced values ranging from the minimum to the maximum values observed for v^2 in the samples drawn is considered. Then, the optimal parameters for these cases are found and the specific parameter value for each sample is obtained by interpolation. This solution appears to be reasonable because of the evidence of a monotone relationship between v^2 and the optimal value for parameter γ or δ.
This chapter is divided into four main parts: first, a deep study of the unconditional quantile Bayes estimators properties is presented (section 4.1); then, the related credible intervals are studied (section 4.2). The simulation study is completed by the analysis of the estimators for the log-normal quantile regression model (section 4.3). Finally, some applications of the methods to real data from different fields are presented (section 4.4).

4.1 Frequentist MSE evaluation

This section contains different comparisons in terms of RRMSE and relative bias among the proposed procedures and the methods which are currently used in the literature. All the Monte Carlo results in this part are obtained with $B = 50,000$ MC replicates.
The section is organized as follows: a comparison of the different priors on ξ is illustrated, with a particular focus on the median case (subsection 4.1.1); then, the effects of introducing a prior guess of σ^2 are investigated (subsection 4.1.2); furthermore, the Bayes estimators are compared to other estimators (subsection 4.1.3). Finally, the efficacy of the posterior variance as an estimate of the frequentist estimator variance (section 4.1.4) and the robustness with respect to model misspecification (section 4.1.5) are studied.

4.1.1 Comparison among different priors on ξ

The first aspect evaluated through an empirical study is the behaviour of the Bayes estimator with different specifications of the priors of the mean in the log-scale ξ. In fact, in this case, the main interest is to check the estimator sensibility with respect to different prior parameters specifications, comparing the normal conjugate prior $\xi|\sigma^2 \sim \mathcal{N}(\xi_0, \sigma^2 n_0^{-1})$ to the flat improper prior $p(\xi) \propto 1$. In particular, n_0 is fixed in order to obtain a prior weight of $1 - \psi = 0.02$, to limit the influence of ξ_0 on the posterior. On the other hand, a grid of values for ξ_0 ranging from -3 to 3 is considered. It is worth to emphasize that, since ξ_0 is related to the log-scale of the data, the considered range includes extremes misspecification scenarios too.

Table 4.1: RRMSE and RB for the estimator of quantile $p = 0.1$. $\hat{\theta}_{0.1}^{QBn}$ is subjected to different prior settings on ξ: flat improper $(-)$ and with the NGIG prior having different values for ξ_0: $(-3, 0, 3)$. $\hat{\theta}^*_{0.1}$ is also reported as a comparison term.

		Relative Root-MSE					Relative-Bias				
	ξ_0:	-	-	-3	0	3	-	-	-3	0	3
σ^2	n	$\hat{\theta}^*_{0.1}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}^*_{0.1}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}_{0.1}^{QBn}$	$\hat{\theta}_{0.1}^{QBn}$
0.25	11	0.042	0.062	0.060	0.060	0.064	-0.006	-0.014	-0.022	-0.005	0.011
	21	0.030	0.043	0.045	0.042	0.046	-0.003	-0.009	-0.021	-0.004	0.011
	51	0.019	0.027	0.031	0.027	0.032	-0.001	-0.003	-0.018	-0.001	0.015
0.5	11	0.035	0.051	0.050	0.051	0.054	-0.007	-0.011	-0.015	-0.005	0.004
	21	0.025	0.036	0.036	0.035	0.037	-0.004	-0.008	-0.014	-0.005	0.005
	51	0.016	0.023	0.024	0.023	0.025	-0.002	-0.003	-0.011	-0.001	0.008
1	11	0.023	0.034	0.033	0.035	0.037	-0.007	-0.007	-0.008	-0.004	0.001
	21	0.017	0.024	0.023	0.024	0.025	-0.004	-0.006	-0.008	-0.004	0.001
	51	0.011	0.013	0.014	0.013	0.014	-0.001	-0.004	-0.007	-0.003	0.001
2	11	0.011	0.017	0.017	0.018	0.019	-0.004	-0.003	-0.003	-0.002	0.000
	21	0.008	0.012	0.011	0.012	0.012	-0.002	-0.004	-0.004	-0.003	-0.001
	51	0.005	0.009	0.009	0.009	0.009	-0.001	-0.004	-0.005	-0.003	-0.002

The only considered estimators are the minimum MSE Bayes estimator, that assumes σ^2 as known ($\hat{\theta}^*_p$), and the Bayes estimators having the hyperparameters fixed by numerically minimizing the MSE as described in section 3.3.2. At the moment, the possibility to incorporate a guess of σ^2 is not considered and the sample variance is plugged into the MSE for the minimization step. Initially, both the Bayes estimator under quadratic loss $\hat{\theta}_p^{QBn}$ and under relative quadratic loss $\hat{\theta}_p^{RQB}$ were included in the Monte Carlo study, but the second resulted to be competitive only in the median case, therefore its results are not reported for other quantiles. Moreover, since hyperparameters specification does not influence the performances of $\hat{\theta}_{0.5}^{RQB}$, the weakly informative setting is chosen.
The behaviour of the estimators is different in the various quantiles. In particular, for quantiles below the median, like $p = 0.10$, whose results are reported in table 4.1, the improvements brought by the NGIG prior are only slight and the frequentist MSE of $\hat{\theta}_{0.10}^{QBn}$ rapidly increase in case of ξ_0 over-specified. The reason why this happens can be explained by the effect that the value of ξ_0 has on the estimator bias: with a correct determination, the MSE is comparable to the one of the estimator with improper prior on ξ, but the reduction of the bias is notable. On the other hand, the underestimation of ξ_0 does not compromise the efficiency of $\hat{\theta}_{0.10}^{QBn}$, since it increases the magnitude of the negative bias; whereas, the overestimation produces an estimator with positive bias and, consequently, inefficient in the log-normal estimation framework (Fabrizi and Trivisano, 2012).

Table 4.2: RRMSE and RB for the estimators of quantile $p = 0.5$. $\hat{\theta}_{0.5}^{QBw}$ and $\hat{\theta}_{0.5}^{RQBw}$ are subjected to different prior settings on ξ: flat improper $(-)$ and with the NGIG prior having different values for ξ_0: $(-3, 0, 3)$. $\hat{\theta}_{0.5}^{*}$ is also reported as a comparison term.

	ξ_0	-	-	-3	0	3	-	-3	0	3
σ^2	n	$\hat{\theta}_{0.5}^{*}$	$\hat{\theta}_{0.5}^{QBw}$	$\hat{\theta}_{0.5}^{QBw}$	$\hat{\theta}_{0.5}^{QBw}$	$\hat{\theta}_{0.5}^{QBw}$	$\hat{\theta}_{0.5}^{RQBw}$	$\hat{\theta}_{0.5}^{RQBw}$	$\hat{\theta}_{0.5}^{RQBw}$	$\hat{\theta}_{0.5}^{RQBw}$
		Relative Root-MSE								
0.25	11	0.079	0.083	0.078	0.081	0.097	0.080	0.085	0.078	0.084
	21	0.057	0.059	0.059	0.057	0.072	0.057	0.064	0.056	0.064
	51	0.037	0.037	0.044	0.037	0.052	0.037	0.047	0.036	0.049
0.5	11	0.086	0.094	0.085	0.092	0.105	0.087	0.088	0.085	0.088
	21	0.062	0.065	0.061	0.064	0.075	0.062	0.066	0.061	0.065
	51	0.040	0.041	0.042	0.040	0.051	0.040	0.046	0.039	0.046
1	11	0.082	0.099	0.087	0.096	0.108	0.084	0.083	0.082	0.084
	21	0.060	0.066	0.059	0.064	0.074	0.060	0.061	0.059	0.061
	51	0.039	0.040	0.038	0.039	0.047	0.039	0.041	0.038	0.042
2	11	0.067	0.095	0.083	0.091	0.100	0.069	0.067	0.067	0.069
	21	0.049	0.059	0.052	0.058	0.065	0.050	0.050	0.049	0.050
	51	0.032	0.035	0.032	0.034	0.039	0.032	0.033	0.031	0.033
		Relative Bias								
0.25	11	-0.012	0.013	-0.019	0.012	0.045	-0.016	-0.044	-0.013	0.018
	21	-0.006	0.006	-0.025	0.006	0.039	-0.007	-0.037	-0.007	0.025
	51	-0.003	0.003	-0.028	0.003	0.035	-0.003	-0.033	-0.003	0.030
0.5	11	-0.018	0.020	-0.006	0.019	0.045	-0.023	-0.042	-0.019	0.004
	21	-0.010	0.010	-0.015	0.009	0.035	-0.011	-0.033	-0.010	0.014
	51	-0.004	0.004	-0.020	0.004	0.029	-0.004	-0.027	-0.004	0.021
1	11	-0.024	0.028	0.008	0.026	0.044	-0.028	-0.039	-0.024	-0.009
	21	-0.013	0.014	-0.004	0.013	0.031	-0.015	-0.029	-0.013	0.003
	51	-0.005	0.006	-0.011	0.005	0.023	-0.006	-0.021	-0.005	0.011
2	11	-0.027	0.033	0.019	0.031	0.043	-0.028	-0.033	-0.025	-0.016
	21	-0.015	0.017	0.005	0.016	0.027	-0.016	-0.023	-0.015	-0.006
	51	-0.006	0.007	-0.003	0.006	0.017	-0.006	-0.015	-0.006	0.003

Observing the quantiles above the median, the implications of the NGIG prior specification are different. In table 4.3 the results of the Monte Carlo study for $p = 0.90$ are reported. In these cases, the optimal estimator is negatively biased, and the NGIG prior is not able to reduce the magnitude of the bias, excluding the case of an important overs-specification of ξ_0. The improvements that are brought by the specified NGIG prior are limited for the

MSE too.

Table 4.3: RRMSE and RB for the estimator of quantile $p = 0.9$. $\hat{\theta}_{0.9}^{QBn}$ is subjected to different prior settings on ξ: flat improper $(-)$ and with the NGIG prior having different values for ξ_0: $(-3, 0, 3)$. $\hat{\theta}^*_{0.9}$ is also reported as a comparison term.

		Relative Root-MSE					Relative-Bias				
	ξ_0:	-	-	-3	0	3	-	-	-3	0	3
σ^2	n	$\hat{\theta}^*_{0.9}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}^*_{0.9}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}_{0.9}^{QBn}$	$\hat{\theta}_{0.9}^{QBn}$
	11	0.151	0.201	0.211	0.197	0.199	-0.023	-0.052	-0.116	-0.063	-0.003
0.25	21	0.109	0.143	0.158	0.141	0.147	-0.012	-0.030	-0.092	-0.038	0.024
	51	0.070	0.093	0.113	0.092	0.104	-0.005	-0.017	-0.075	-0.022	0.042
	11	0.212	0.271	0.279	0.267	0.266	-0.045	-0.088	-0.155	-0.104	-0.047
0.5	21	0.153	0.198	0.211	0.196	0.197	-0.024	-0.053	-0.118	-0.065	-0.005
	51	0.099	0.131	0.148	0.130	0.135	-0.010	-0.030	-0.090	-0.036	0.026
	11	0.296	0.363	0.368	0.358	0.355	-0.088	-0.148	-0.216	-0.168	-0.116
1	21	0.216	0.273	0.283	0.271	0.269	-0.047	-0.091	-0.158	-0.107	-0.051
	51	0.140	0.182	0.197	0.181	0.182	-0.019	-0.050	-0.112	-0.059	0.001
	11	0.410	0.476	0.479	0.472	0.468	-0.167	-0.245	-0.310	-0.268	-0.222
2	21	0.301	0.371	0.377	0.368	0.365	-0.092	-0.153	-0.221	-0.173	-0.121
	51	0.196	0.253	0.264	0.251	0.250	-0.038	-0.083	-0.147	-0.095	-0.038

A very particular case is represented by the median. In table 4.2 the outputs related both to the estimator under quadratic loss $\hat{\theta}_{0.5}^{QBw}$ and the estimator under relative quadratic loss $\hat{\theta}_{0.5}^{RQBw}$ with the weakly informative prior setting are reported. As expected from the findings reported in section 3.3.2, in the case of improper prior on ξ the posterior mean is not an efficient estimation because of its positive bias. On the other hand, the estimator under relative quadratic loss reaches the frequentist MSE of the minimum MSE conditioned estimator. This is a strong result, caused by the fact that σ^2 does not enter the functional to estimate.

Furthermore, in this case, the NGIG prior results to be convenient for both the estimators. As far $\hat{\theta}_{0.5}^{QBw}$ it concerns, a proper prior on ξ induces an acceptable estimator, unless a strong over-specification of ξ_0 is produces, whereas, $\hat{\theta}_{0.5}^{RQBw}$ results to be robust to extreme misspecifications of the prior parameter ξ_0. Moreover, for a wide set of values of ξ_0, the Bayes estimator results to be more efficient than the conditioned estimator $\hat{\theta}^*_{0.5}$, which is based on the flat improper prior on ξ.

The characteristics of the estimators described above can be verified in figure 4.1 too, where CRMSE and relative bias are shown. It is reported the scenario with $\sigma^2 = 0.5$ and a sample size of 11 or 21. It well represents the behaviour of the estimators with respect to the hyperparameter ξ_0 variation.

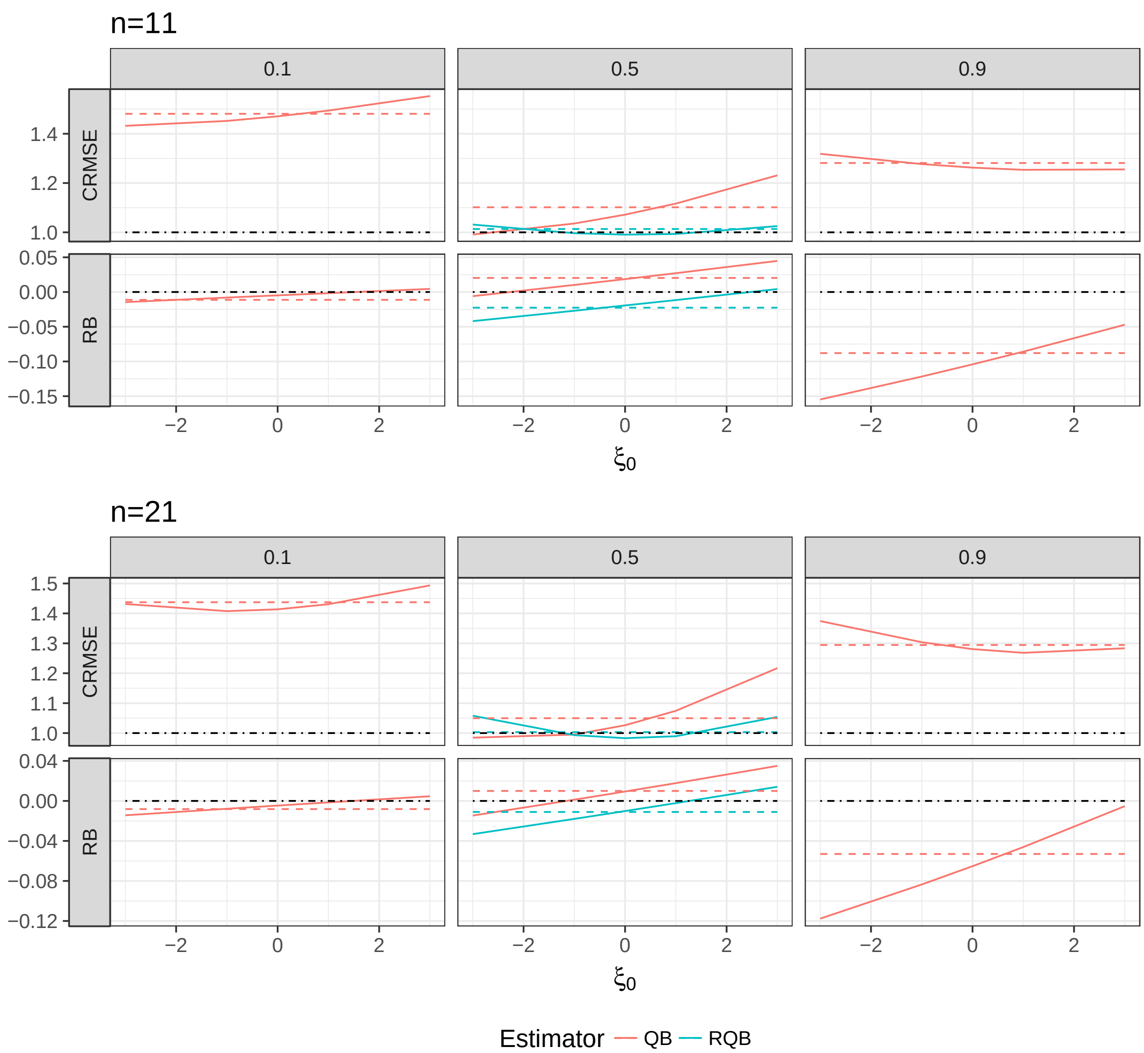

Figure 4.1: Comparison between the flat improper prior on ξ (dashed lines) and the NGIG prior (solid lines). The changes of the CRMSE and relative bias with respect to ξ_0 are shown for the scenarios with $n = 11$ and $n = 21$ with $\sigma^2 = 0.5$. The results for quantiles 0.1, 0.5 and 0.9 are reported.

4.1.2 Methods incorporating a guess on σ^2

A further aim of the simulation study is to investigate the consequences of setting a guess s_0^2 for σ^2 in the procedures to deduce the optimal prior parameters. A set of values for s_0^2, determined as a portion of the true σ^2 (from $0.4\sigma^2$ to $2\sigma^2$), is used in the simulation.
When the numerically optimal prior is adopted, the procedure of plugging the sample estimate of σ^2 in the MSE formula has two main drawbacks: the double use of data both in

			Relative Root MSE				Relative Bias			
		s_0^2:	-	$0.4\sigma^2$	σ^2	$2\sigma^2$	-	$0.4\sigma^2$	σ^2	$2\sigma^2$
p	σ^2	n	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$
0.05	0.25	11	0.059	0.053	0.050	0.060	-0.013	-0.001	-0.017	-0.048
		21	0.041	0.039	0.037	0.042	-0.009	-0.001	-0.010	-0.029
		51	0.026	0.025	0.025	0.025	-0.003	0.000	0.001	0.001
	0.5	11	0.045	0.041	0.036	0.048	-0.010	0.003	-0.014	-0.042
		21	0.032	0.030	0.028	0.034	-0.008	0.001	-0.009	-0.028
		51	0.020	0.019	0.020	0.025	-0.003	-0.001	0.002	-0.020
	1	11	0.027	0.025	0.020	0.028	-0.005	0.004	-0.009	-0.027
		21	0.019	0.018	0.016	0.025	-0.005	0.002	-0.006	-0.023
		51	0.012	0.011	0.011	0.019	-0.001	-0.001	-0.005	-0.017
	2	11	0.012	0.012	0.007	0.012	-0.002	0.004	-0.004	-0.011
		21	0.008	0.008	0.006	0.008	-0.002	0.001	-0.004	-0.007
		51	0.007	0.004	0.005	0.011	-0.002	-0.001	-0.003	-0.011
0.25	0.25	11	0.066	0.063	0.061	0.069	-0.012	0.001	-0.014	-0.045
		21	0.047	0.045	0.045	0.048	-0.008	0.000	-0.008	-0.025
		51	0.030	0.029	0.029	0.029	-0.003	0.000	0.001	0.001
	0.5	11	0.063	0.060	0.056	0.069	-0.013	0.004	-0.015	-0.053
		21	0.043	0.043	0.042	0.047	-0.006	0.001	-0.009	-0.030
		51	0.028	0.027	0.028	0.028	-0.002	0.000	0.002	-0.010
	1	11	0.051	0.049	0.043	0.058	-0.012	0.007	-0.015	-0.050
		21	0.035	0.035	0.032	0.044	-0.008	0.003	-0.008	-0.037
		51	0.021	0.022	0.022	0.022	-0.003	0.000	-0.002	0.000
	2	11	0.032	0.034	0.026	0.037	-0.009	0.008	-0.011	-0.035
		21	0.022	0.023	0.020	0.027	-0.006	0.003	-0.009	-0.023
		51	0.016	0.013	0.014	0.027	-0.006	-0.001	0.002	-0.026

Table 4.4: RRMSE and RB for Bayes estimators of quantiles 0.05 and 0.25. A guess s_0^2 of σ^2 is included in the choice of the hyperparameters.

			Relative Root MSE				Relative Bias			
		s_0^2:	-	$0.4\sigma^2$	σ^2	$2\sigma^2$	-	$0.4\sigma^2$	σ^2	$2\sigma^2$
p	σ^2	n	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$	$\hat{\theta}_p^{QBn}$
0.75	0.25	11	0.121	0.122	0.114	0.116	-0.028	-0.065	-0.021	0.002
		21	0.088	0.093	0.084	0.086	-0.015	-0.053	-0.014	0.004
		51	0.057	0.063	0.055	0.056	-0.008	-0.036	-0.008	0.003
	0.5	11	0.147	0.147	0.142	0.143	-0.043	-0.073	-0.037	-0.022
		21	0.108	0.111	0.105	0.106	-0.024	-0.057	-0.023	-0.008
		51	0.070	0.074	0.069	0.070	-0.012	-0.037	-0.012	-0.002
	1	11	0.169	0.168	0.166	0.166	-0.062	-0.081	-0.057	-0.053
		21	0.126	0.127	0.123	0.124	-0.036	-0.063	-0.035	-0.022
		51	0.083	0.085	0.081	0.082	-0.017	-0.041	-0.018	-0.007
	2	11	0.178	0.178	0.177	0.178	-0.085	-0.089	-0.080	-0.089
		21	0.134	0.135	0.133	0.133	-0.049	-0.069	-0.049	-0.041
		51	0.090	0.091	0.089	0.102	-0.024	-0.045	-0.024	0.023
0.95	0.25	11	0.274	0.282	0.206	0.239	-0.072	-0.234	-0.053	0.059
		21	0.194	0.229	0.163	0.183	-0.042	-0.193	-0.041	0.036
		51	0.126	0.162	0.115	0.123	-0.026	-0.134	-0.026	0.015
	0.5	11	0.395	0.390	0.326	0.350	-0.130	-0.310	-0.117	-0.006
		21	0.289	0.314	0.255	0.273	-0.082	-0.249	-0.084	0.001
		51	0.191	0.219	0.178	0.185	-0.048	-0.168	-0.049	-0.001
	1	11	0.581	0.564	0.512	0.532	-0.241	-0.438	-0.237	-0.119
		21	0.440	0.455	0.401	0.420	-0.155	-0.351	-0.163	-0.057
		51	0.296	0.321	0.279	0.289	-0.089	-0.235	-0.092	-0.025
	2	11	0.874	0.851	0.809	0.826	-0.457	-0.671	-0.461	-0.341
		21	0.692	0.696	0.638	0.662	-0.298	-0.546	-0.314	-0.170
		51	0.475	0.502	0.450	0.466	-0.170	-0.371	-0.177	-0.071

Table 4.5: Root MSE and relative bias for Bayes estimators of quantile 0.75 and 0.95. A guess s_0^2 of σ^2 is included in the choice of the hyperparameters.

the prior specification and in the likelihood, and the increase of the estimator MSE caused by the sample variability of the estimate v^2.
Therefore, the performances of the Bayes estimators are compared in order to understand the robustness with respect to the misspecification of s_0^2 and to quantify the gains in efficiency of including a guess of σ^2 instead of v^2
The results about quantiles 0.05, 0.25, 0.75 and 0.95 are reported in table 4.4 and table 4.5; but a general consideration can be inferred in this case. It is evident that, considering $\hat{\theta}_p^{QBn}$, the gain in terms of efficiency obtained plugging s_0^2 into the MSE occurs in all the considered scenarios, even if they are more marked with small samples, as would be expected. Moreover, a moderate robustness with respect to the misspecification of s_0^2 can be noted. Therefore, if in a particular inferential problem a reliable value for σ^2 can be hypothesised, the Bayes estimator with the optimal hyperparameters chosen minimizing the MSE with s_0^2 instead of v^2 results to be a safe and efficient estimator for all the quantiles. Finally, it must be stressed that the improvements in the MSE do not increase the negative bias. It is worth to remember that this procedure cannot be considered for the median, where the optimization algorithm is not used in the hyperparameters choice.

4.1.3 General comparison

Another goal of the simulation study is the comparison of the proposed Bayes estimators to the methods which are already present in the literature, in particular the Longford's minimum MSE estimator $\hat{Q}_p$, the naive estimator $\tilde{\theta}_p$, and, as benchmark, the Bayes estimator with minimum MSE $\hat{\theta}_p^*$ conditioned on σ^2. To enrich the simulation study, the `R` default non-parametric estimator $\hat{Q}_p^7$ defined in section 3.1.1 is considered too. An exhaustive grid of quantiles, with a particular focus on the extremes, and different combinations of σ^2 and n are considered. In tables 4.6, 4.7, D.1 and D.2 the RRMSE and the relative bias of different estimators are reported for different quantiles and sample sizes, with σ^2 respectively equal to: 0.25, 1, 0.5, 2. Moreover, the behaviour of the estimators performances is shown in figure 4.2 for the case $\sigma^2 = 0.5$. In the median case, from now on, with $\hat{\theta}_p^{Bn}$ the Bayes estimator under relative quadratic with the weakly informative prior $\hat{\theta}_p^{RQBw}$ is intended, whereas, in the other cases, the flat improper prior is adopted for ξ and v^2 is plugged in the MSE to minimize. The Bayes estimator $\hat{\theta}_p^{QBa}$, with the approximately optimal priors cannot be evaluated in the median case, as explained in section 3.3.
By looking at the RRMSE, the Bayes estimators outperform almost everywhere the other considered estimators. Considerable improvements in the RRMSE are evident with small sample sizes and extreme quantiles: in these cases $\hat{\theta}_p^{QBn}$ always has the lowest RMSE than the others when $p > 0.90$, followed by $\hat{\theta}_p^{QBa}$. In the smaller quantiles, the estimators reported quite similar RMSEs, with the exception of the naive estimator that always presents the worst performance among the parametric estimators. Another interesting point to highlight is about the median estimation: since in this case the estimand does not involve σ^2 the Bayes conditioned estimator $\hat{\theta}_p^*$ is reached by the other competitors.
The non-parametric estimator is not efficient if compared to the model based estimators,

		Relative Root MSE								
		p								
n	Method	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
11	$\hat{\theta}_p^*$	0.025	0.035	0.042	0.057	0.079	0.111	0.151	0.181	0.254
	$\hat{Q}_p^7$	0.112	0.082	0.082	0.079	0.101	0.140	0.231	0.301	0.615
	$\hat{\theta}_p$	0.054	0.059	0.061	0.066	0.081	0.126	0.212	0.294	0.536
	$\hat{Q}_p$	0.052	0.058	0.061	0.066	**0.080**	**0.120**	0.202	0.282	0.527
	$\hat{\theta}_p^{Ba}$	**0.049**	**0.055**	**0.058**	**0.064**	-	0.123	**0.200**	**0.273**	0.482
	$\hat{\theta}_p^{Bn}$	0.053	0.059	0.062	0.066	**0.080**	0.121	0.201	0.274	**0.480**
21	$\hat{\theta}_p^*$	0.018	0.025	0.030	0.041	0.057	0.080	0.109	0.130	0.183
	$\hat{Q}_p^7$	0.075	0.059	0.056	0.058	0.073	0.108	0.175	0.247	0.490
	$\hat{\theta}_p$	0.037	0.041	0.042	**0.046**	0.058	0.090	0.151	0.207	0.371
	$\hat{Q}_p$	0.036	0.040	0.042	0.047	**0.057**	**0.088**	0.147	0.204	0.371
	$\hat{\theta}_p^{Ba}$	**0.035**	**0.039**	**0.041**	**0.046**	-	0.089	0.146	0.199	0.353
	$\hat{\theta}_p^{Bn}$	0.037	0.041	0.043	0.047	**0.057**	**0.088**	**0.143**	**0.194**	**0.338**
51	$\hat{\theta}_p^*$	0.012	0.016	0.019	0.026	0.037	0.052	0.070	0.084	0.118
	$\hat{Q}_p^7$	0.044	0.035	0.034	0.036	0.046	0.069	0.117	0.163	0.344
	$\hat{\theta}_p$	0.023	**0.025**	0.027	**0.029**	**0.037**	0.058	0.095	0.130	0.231
	$\hat{Q}_p$	0.023	**0.025**	0.027	**0.029**	**0.037**	**0.057**	0.094	0.129	0.231
	$\hat{\theta}_p^{Ba}$	**0.022**	**0.025**	**0.026**	**0.029**	-	**0.057**	0.094	0.128	0.226
	$\hat{\theta}_p^{Bn}$	0.023	0.026	0.027	0.030	**0.037**	**0.057**	**0.093**	**0.126**	**0.221**
		Relative Bias								
11	$\hat{\theta}_p^*$	-0.004	-0.005	-0.006	-0.009	-0.012	-0.017	-0.023	-0.027	-0.039
	$\hat{Q}_p^7$	0.090	0.051	0.040	0.023	0.009	-0.013	-0.080	-0.131	-0.499
	$\hat{\theta}_p$	0.012	0.011	0.010	0.008	0.006	0.004	0.006	0.010	0.029
	$\hat{Q}_p$	-0.003	-0.006	-0.007	-0.010	-0.013	-0.012	-0.005	0.004	0.041
	$\hat{\theta}_p^{Ba}$	-0.003	-0.005	-0.005	-0.001	-	-0.014	-0.046	-0.070	-0.125
	$\hat{\theta}_p^{Bn}$	-0.012	-0.013	-0.014	-0.012	-0.016	-0.028	-0.052	-0.072	-0.122
21	$\hat{\theta}_p^*$	-0.002	-0.003	-0.003	-0.005	-0.006	-0.009	-0.012	-0.015	-0.021
	$\hat{Q}_p^7$	0.057	0.030	0.022	0.012	0.005	-0.008	-0.045	-0.100	-0.345
	$\hat{\theta}_p$	0.006	0.006	0.005	0.004	0.003	0.002	0.003	0.005	0.015
	$\hat{Q}_p$	-0.002	-0.003	-0.004	-0.006	-0.007	-0.006	-0.001	0.004	0.026
	$\hat{\theta}_p^{Ba}$	-0.002	-0.003	-0.003	-0.001	-	-0.007	-0.024	-0.037	-0.066
	$\hat{\theta}_p^{Bn}$	-0.009	-0.009	-0.009	-0.008	-0.007	-0.015	-0.030	-0.042	-0.078
51	$\hat{\theta}_p^*$	-0.001	-0.001	-0.001	-0.002	-0.003	-0.004	-0.005	-0.006	-0.008
	$\hat{Q}_p^7$	0.028	0.013	0.009	0.005	0.002	-0.003	-0.019	-0.041	-0.172
	$\hat{\theta}_p$	0.003	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.005
	$\hat{Q}_p$	-0.001	-0.001	-0.002	-0.002	-0.003	-0.002	0.000	0.002	0.010
	$\hat{\theta}_p^{Ba}$	-0.001	-0.001	-0.001	0.000	-	-0.003	-0.010	-0.016	-0.028
	$\hat{\theta}_p^{Bn}$	-0.003	-0.003	-0.003	-0.003	-0.003	-0.008	-0.017	-0.026	-0.050

Table 4.6: RRMSE and RB of estimators for θ_p with respect to different sample sizes n and quantiles p, with $\sigma^2 = 0.25$.

		Relative Root MSE								
		p								
n	Method	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
11	$\hat{\theta}_p^*$	0.008	0.016	0.023	0.042	0.082	0.161	0.296	0.426	0.842
	$\hat{Q}_p^7$	0.057	0.051	0.056	0.069	0.116	0.222	0.468	0.773	1.708
	$\hat{\theta}_p$	0.022	0.031	0.038	0.054	0.091	0.199	0.470	0.807	2.262
	$\hat{Q}_p$	0.020	0.028	0.035	**0.051**	**0.084**	0.169	0.379	0.647	1.767
	$\hat{\theta}_p^{Ba}$	0.020	0.028	0.035	0.052	-	0.190	0.416	0.690	1.874
	$\hat{\theta}_p^{Bn}$	**0.018**	**0.027**	**0.034**	**0.051**	**0.084**	**0.169**	**0.363**	**0.581**	**1.394**
21	$\hat{\theta}_p^*$	0.006	0.012	0.017	0.030	0.060	0.117	0.216	0.310	0.613
	$\hat{Q}_p^7$	0.034	0.033	0.035	0.047	0.080	0.164	0.355	0.582	1.499
	$\hat{\theta}_p$	0.014	0.020	0.025	0.036	0.063	0.138	0.316	0.530	1.396
	$\hat{Q}_p$	0.013	**0.019**	**0.024**	**0.035**	**0.060**	**0.126**	0.286	0.481	1.287
	$\hat{\theta}_p^{Ba}$	0.013	**0.019**	**0.024**	0.036	-	0.135	0.301	0.497	1.300
	$\hat{\theta}_p^{Bn}$	**0.012**	**0.019**	**0.024**	**0.035**	**0.060**	**0.126**	**0.273**	**0.440**	**1.067**
51	$\hat{\theta}_p^*$	0.004	0.007	0.011	0.020	0.039	0.076	0.140	0.201	0.397
	$\hat{Q}_p^7$	0.018	0.018	0.020	0.028	0.050	0.103	0.234	0.392	1.173
	$\hat{\theta}_p$	0.008	**0.012**	0.015	0.023	0.040	0.086	0.194	0.320	0.812
	$\hat{Q}_p$	0.008	**0.012**	0.015	0.022	**0.039**	**0.083**	0.187	0.309	0.792
	$\hat{\theta}_p^{Ba}$	0.008	**0.012**	0.015	0.022	-	0.086	0.191	0.313	0.793
	$\hat{\theta}_p^{Bn}$	**0.007**	**0.012**	**0.013**	**0.021**	**0.039**	**0.083**	**0.182**	**0.296**	**0.721**
		Relative Bias								
11	$\hat{\theta}_p^*$	-0.002	-0.005	-0.007	-0.012	-0.024	-0.048	-0.088	-0.126	-0.249
	$\hat{Q}_p^7$	0.043	0.032	0.029	0.025	0.020	0.006	-0.106	-0.177	-1.288
	$\hat{\theta}_p$	0.007	0.009	0.009	0.011	0.013	0.022	0.057	0.110	0.387
	$\hat{Q}_p$	-0.001	-0.002	-0.005	-0.012	-0.025	-0.043	-0.051	-0.038	0.084
	$\hat{\theta}_p^{Ba}$	0.005	0.006	0.007	0.011	-	0.002	-0.034	-0.060	-0.053
	$\hat{\theta}_p^{Bn}$	-0.002	-0.005	-0.007	-0.012	-0.028	-0.062	-0.148	-0.241	-0.579
21	$\hat{\theta}_p^*$	-0.001	-0.003	-0.004	-0.007	-0.013	-0.026	-0.047	-0.068	-0.134
	$\hat{Q}_p^7$	0.025	0.018	0.015	0.013	0.010	-0.001	-0.059	-0.179	-0.890
	$\hat{\theta}_p$	0.003	0.004	0.005	0.005	0.006	0.011	0.029	0.055	0.187
	$\hat{Q}_p$	0.000	-0.002	-0.003	-0.007	-0.014	-0.021	-0.021	-0.008	0.091
	$\hat{\theta}_p^{Ba}$	0.002	0.002	0.003	0.004	-	0.003	-0.013	-0.021	0.001
	$\hat{\theta}_p^{Bn}$	-0.003	-0.005	-0.006	-0.008	-0.015	-0.036	-0.091	-0.155	-0.400
51	$\hat{\theta}_p^*$	-0.001	-0.001	-0.001	-0.003	-0.005	-0.011	-0.019	-0.028	-0.055
	$\hat{Q}_p^7$	0.012	0.008	0.006	0.005	0.004	0.000	-0.025	-0.067	-0.409
	$\hat{\theta}_p$	0.001	0.002	0.002	0.002	0.003	0.005	0.011	0.021	0.069
	$\hat{Q}_p$	0.000	-0.001	-0.001	-0.003	-0.005	-0.008	-0.007	-0.001	0.041
	$\hat{\theta}_p^{Ba}$	0.001	0.001	0.001	0.002	-	0.002	-0.004	-0.007	0.002
	$\hat{\theta}_p^{Bn}$	-0.003	-0.001	-0.004	-0.003	-0.006	-0.017	-0.050	-0.089	-0.244

Table 4.7: RRMSE and RB of estimators for θ_p with respect to different sample sizes n and quantiles p, with $\sigma^2 = 1$. The Bayes estimator $\hat{\theta}_p^B$ is the estimator under relative quadratic loss for the median and the one under quadratic loss for the others.

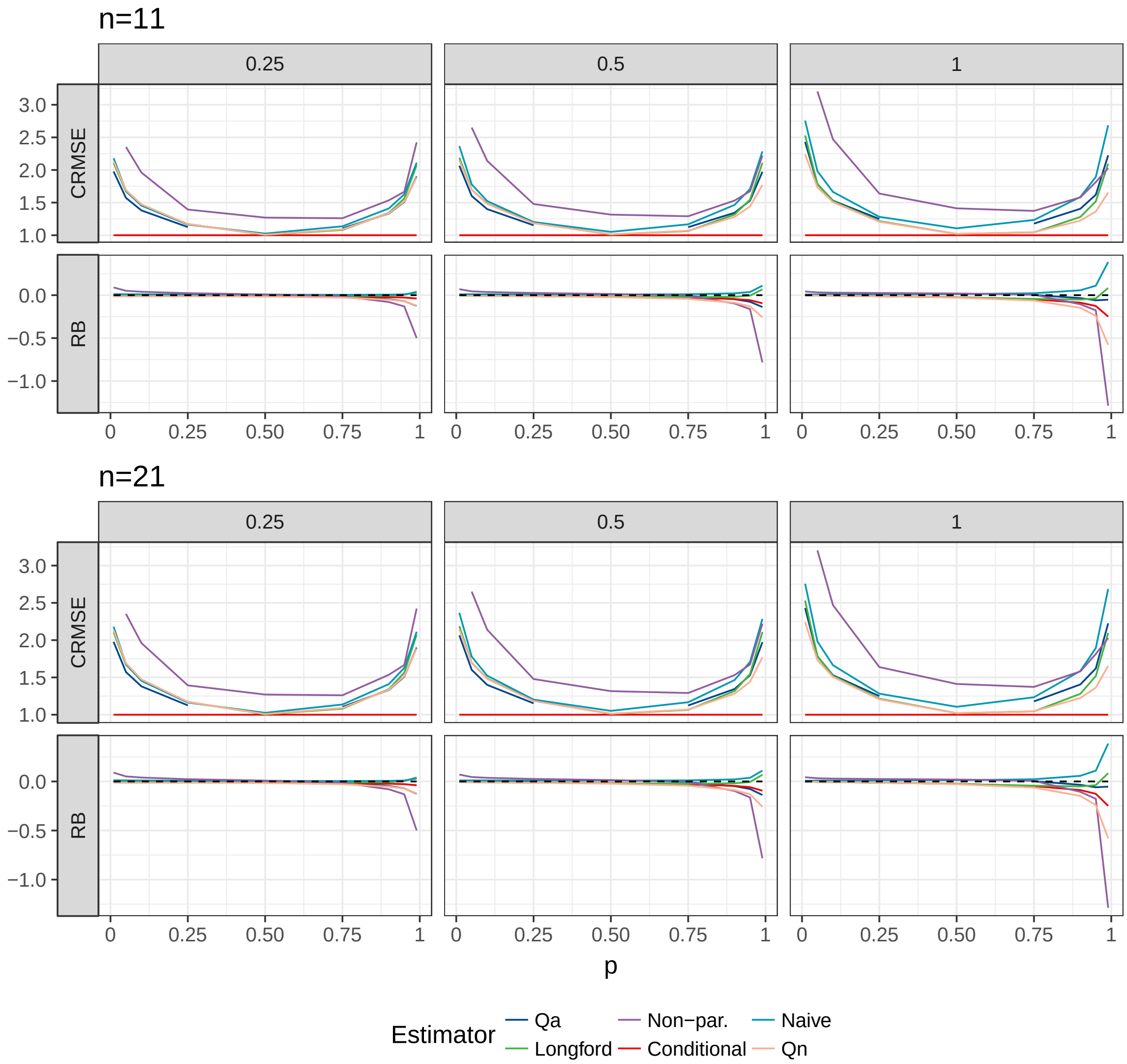

Figure 4.2: Comparison at the different quantile levels among the considered estimators in terms of CRMSE and RB. For the sample sizes of 11 and 21 the variance is fixed equal to 0.25, 0.5, 1.

in particular when the variance in the log-scale increases. Moreover, the major critical points are the extreme quantiles, since this particular estimator assumes the lowest sample observation as $p = 0$ and the bigger as $p = 1$. This fact implies a considerable positive bias in the estimation of the lower quantiles and a negative one in estimating the higher quantiles, even if in the latter case the RRMSE is limited.
It might be noted that the Bayes estimators, $\hat{\theta}_p^{Bn}$ mainly, limit the RRMSE through an important presence of negative bias. This is true also for the conditioned estimator $\hat{\theta}_p^{*}$. In

particular, $\hat{\theta}_p^{QBn}$ appears to be often the most biased estimator, whereas $\hat{\theta}_p^{QBa}$ keeps the bias low.
A negative aspect of the main competitor $\hat{Q}_p$ is underlined in this simulation scheme: some computational issues are encountered for low quantiles with particular combinations of σ^2 and n because of the existence conditions reported in section 3.1.3.
Finally, comparing tables 4.6 and 4.7 it can be noted that, in estimating quantiles below the median, the RRMSE decreases when σ^2 increases. A possible interpretation of this behaviour is connected to the distribution skewness (1.8), which rapidly increases with σ^2. Moreover, at the same time, the mode of the distribution, equal to $\exp\left\{\xi - \sigma^2\right\}$, approaches to 0. Consequently, in that cases, the amount of information contained in the left tail of the distribution is higher, leading to a more precise quantile estimate.

4.1.4 Posterior variance

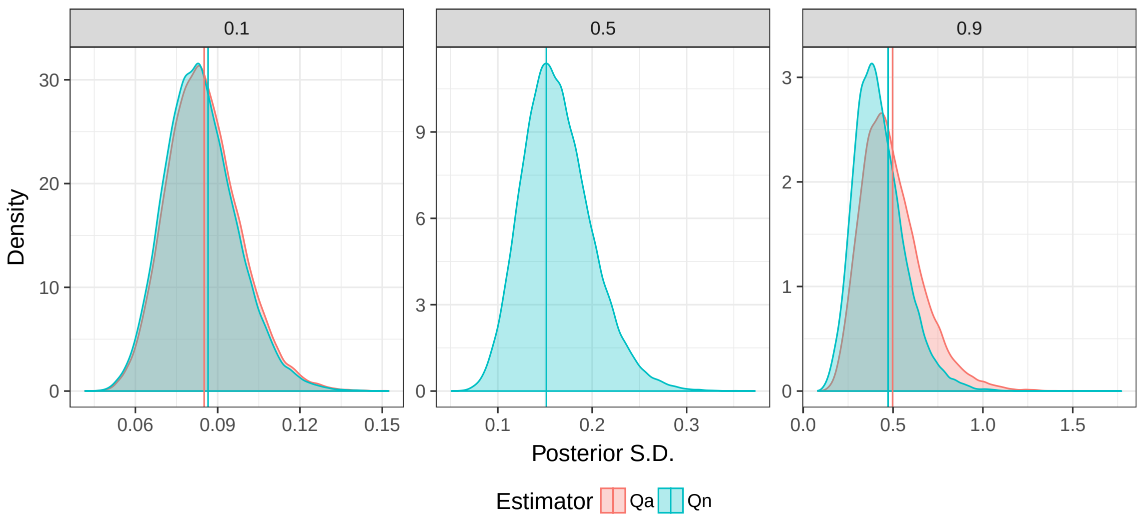

Figure 4.3: Kernel density plots of the posterior standard deviation distributions of the $\hat{\theta}_p^{Bn}$ and $\hat{\theta}_p^{QBa}$ estimators at different quantiles p. The case $\sigma^2 = 0.5$ and $n = 21$ is reported. The vertical lines show the square root of the Monte Carlo variances averages of the estimators

Another property assessed through a Monte Carlo study is the reliability of the variance of the θ_p posterior distribution as an estimate of the frequentist variance of estimates. It is an important aspect in order to provide an uncertainty measure to the estimator. The Monte Carlo standard deviation $\sqrt{V_{MC}}$ of the point estimators and the average posterior standard deviation $\sqrt{\bar{V}_{QB}}$ are considered. The results for both the proposed Bayes estimators $\hat{\theta}_p^{Bn}$ and $\hat{\theta}_p^{QBa}$ are reported in table 4.8.

Table 4.8: Monte Carlo standard deviation of the estimator $\hat{\theta}_p^{Bn}$ in the sample space and square root of the expectation of the posterior variance distribution at different n, σ^2, p.

		p:	0.10		0.25		0.50		0.75		0.90	
	σ^2	n	$\sqrt{V_{MC}}$	$\sqrt{\bar{V}_{QB}}$	$\sqrt{V_{MC}}$	$\sqrt{\bar{V}_{QB}}$	$\sqrt{V_{MC}}$	$\sqrt{\bar{V}_{RQB}}$	$\sqrt{V_{MC}}$	$\sqrt{\bar{V}_{QB}}$	$\sqrt{V_{MC}}$	$\sqrt{\bar{V}_{QB}}$
$\hat{\theta}_p^{Bn}$	0.25	11	0.113	0.111	0.123	0.137	0.149	0.171	0.225	0.188	0.373	0.316
		21	0.080	0.080	0.088	0.094	0.108	0.117	0.164	0.150	0.266	0.245
		51	0.051	0.051	0.056	0.057	0.070	0.072	0.110	0.104	0.174	0.166
	0.5	11	0.123	0.116	0.152	0.170	0.208	0.246	0.349	0.287	0.639	0.572
		21	0.087	0.085	0.108	0.116	0.151	0.167	0.261	0.238	0.473	0.448
		51	0.057	0.054	0.070	0.070	0.098	0.102	0.186	0.168	0.351	0.325
	1	11	0.122	0.111	0.178	0.207	0.287	0.363	0.565	0.440	1.192	1.108
		21	0.084	0.081	0.124	0.136	0.211	0.243	0.433	0.386	0.924	0.897
		51	0.046	0.053	0.075	0.082	0.138	0.147	0.297	0.267	0.631	0.619
	2	11	0.108	0.097	0.192	0.257	0.393	0.565	0.956	0.679	2.483	2.363
		21	0.069	0.066	0.131	0.149	0.290	0.364	0.765	0.663	2.052	2.046
		51	0.052	0.041	0.092	0.089	0.193	0.213	0.540	0.500	1.426	1.543
$\hat{\theta}_p^{QBa}$	0.25	11	0.108	0.113	0.121	0.130	-	-	0.229	0.222	0.361	0.372
		21	0.078	0.080	0.087	0.091	-	-	0.168	0.165	0.271	0.276
		51	0.050	0.051	0.056	0.057	-	-	0.108	0.108	0.177	0.178
	0.5	11	0.119	0.120	0.152	0.158	-	-	0.374	0.369	0.658	0.709
		21	0.085	0.086	0.108	0.111	-	-	0.273	0.273	0.498	0.522
		51	0.054	0.055	0.069	0.070	-	-	0.176	0.177	0.327	0.333
	1	11	0.126	0.121	0.188	0.189	-	-	0.652	0.659	1.328	1.515
		21	0.086	0.085	0.129	0.131	-	-	0.474	0.483	1.018	1.113
		51	0.053	0.053	0.080	0.081	-	-	0.305	0.309	0.673	0.701
	2	11	0.129	0.119	0.236	0.228	-	-	1.250	1.307	3.092	3.790
		21	0.079	0.077	0.148	0.148	-	-	0.898	0.941	2.412	2.794
		51	0.046	0.046	0.089	0.089	-	-	0.576	0.590	1.620	1.738

With smaller σ^2, the posterior variance for the prior setting of $\hat{\theta}_p^{QBn}$ tends to underestimate the the estimator variance $\sqrt{V_{MC}}$ especially in the case $n = 11$; whereas it results to be more precise when $\sigma^2 \geq 1$. However, already with $n = 21$ the differences are limited.
The other proposal $\hat{\theta}_p^{QBa}$ shows an opposite behaviour: overestimation is present at the upper quantiles when $\sigma^2 \leq 1$.
On the other hand, the posterior variance with the weakly informative prior, in the median case, tends to overestimate the variance of the Bayes estimator under relative quadratic loss. To compare the Monte Carlo estimate of the estimator variance with respect to the distribution of the posterior variances in the sample space, figure 4.3 is produced.

4.1.5 Robustness with respect to model misspecification

A further aspect that is assessed through simulations is the robustness of the estimators in case of model misspecification. The Bayes estimators, the naive estimator, the Longford's estimator and the non-parametric estimator are included in the analysis. Then, samples from gamma and Weibull distributions are generated. The simulation results are based on $B = 50,000$ replicates and RRMSE and relative bias are reported for different parameters scenarios (table 4.9).
Both the proposed estimators have similar frequentist properties of the estimator by Longford. The main critical points to highlight concern $\hat{\theta}_p^{Qn}$ that possess the higher RRMSE in the right tail quantiles with small sample size ($n = 11$).
Different results are obtained for the median estimation: in this case the naive estimator results to be almost always the most efficient choice, whereas the Bayes estimator outperforms the non-parametric estimator only when the true model is a gamma.
A final observation is about the fact that the quote of RRMSE and RB of each estimator is caused by the kind of distribution, whereas the parameters do not change it.

4.2 Interval estimation: frequentist coverage

A simulation study is also performed to evaluate the interval estimation: the naive confidence intervals obtained exponentiating the limits for the normal quantiles (see section 3.1.2) are compared to the Bayesian credible intervals based on the quantiles of the θ_p posterior distribution in the weakly informative prior setting. The other hyperparameters selection procedures do not result to be appropriated for this task because the related posterior distributions are too peaked and the intervals would not reach the nominal coverage. An example of this fact is reported in figure 4.4 where the results of the intervals based on the prior of $\hat{\theta}_p^{Qn}$ are displayed too.
In table 4.10 the output of the Monte Carlo study based on $B = 10,000$ replications is reported. As it is possible to notice (from figure 4.4 too), the nominal coverage, fixed equal to 0.95, is reached by both the procedures and no critical situations are evidenced. Moreover, an interesting result is about the average width of the intervals: especially in extreme scenarios (i.e. n small and high p), the Bayesian credible intervals are always consistently narrower

Table 4.9: Performances of different estimators of the target functional θ_p in case of misspecified log-normal distribution. Different generating distribution and various samples sizes are considered.

			Relative Root MSE					Relative bias				
p		n	$\hat{\theta}_p$	$\hat{Q}_p$	$\hat{\theta}_p^{QBa}$	$\hat{\theta}_p^{QBn}$	$\hat{Q}_p^7$	$\hat{\theta}_p$	$\hat{Q}_p$	$\hat{\theta}_p^{QBa}$	$\hat{\theta}_p^{QBn}$	$\hat{Q}_p^7$
	Gamma	11	0.501	0.466	0.455	0.463	0.672	0.162	0.030	0.107	0.026	0.354
	sh.=2	21	0.348	0.332	0.329	0.328	0.445	0.097	0.026	0.065	0.007	0.190
	sc.=2	51	0.218	0.212	0.212	0.228	0.267	0.054	0.025	0.041	-0.021	0.133
	Gamma	11	0.501	0.466	0.455	0.463	0.685	0.162	0.030	0.107	0.026	0.354
	sh.=2	21	0.348	0.332	0.329	0.328	0.450	0.097	0.026	0.065	0.007	0.190
0.1	sc.=0.5	51	0.218	0.212	0.212	0.228	0.268	0.054	0.025	0.041	-0.021	0.079
	Weibull	11	0.415	0.402	0.381	0.399	0.530	0.122	0.032	0.064	0.023	0.263
	sh.=2	21	0.298	0.291	0.284	0.291	0.367	0.074	0.026	0.043	0.008	0.145
	sc.=2	51	0.191	0.188	0.187	0.177	0.228	0.041	0.021	0.028	0.005	0.062
	Weibull	11	0.415	0.402	0.381	0.399	0.530	0.122	0.032	0.064	0.023	0.263
	sh.=2	21	0.298	0.291	0.284	0.291	0.367	0.074	0.026	0.043	0.008	0.145
	sc.=0.5	51	0.191	0.188	0.187	0.177	0.228	0.041	0.021	0.028	0.005	0.062
	Gamma	11	0.232	0.264	-	0.268	0.289	-0.062	-0.139	-	-0.148	0.029
	sh.=2	21	0.178	0.199	-	0.201	0.209	-0.077	-0.118	-	-0.121	0.014
	sc.=2	51	0.133	0.145	-	0.145	0.133	-0.085	-0.103	-	-0.103	0.006
	Gamma	11	0.232	0.264	-	0.268	0.289	-0.062	-0.139	-	-0.148	0.029
	sh.=2	21	0.178	0.199	-	0.201	0.209	-0.077	-0.118	-	-0.121	0.014
0.5	sc.=0.5	51	0.133	0.145	-	0.145	0.133	-0.085	-0.103	-	-0.103	0.006
	Weibull	11	0.191	0.222	-	0.226	0.211	-0.082	-0.130	-	-0.137	0.010
	sh.=2	21	0.154	0.173	-	0.174	0.155	-0.090	-0.116	-	-0.118	0.006
	sc.=2	51	0.126	0.135	-	0.135	0.101	-0.096	-0.107	-	-0.107	0.002
	Weibull	11	0.193	0.225	-	0.229	0.213	-0.084	-0.133	-	-0.140	0.008
	sh.=2	21	0.155	0.174	-	0.175	0.155	-0.091	-0.117	-	-0.119	0.006
	sc.=0.5	51	0.126	0.135	-	0.135	0.101	-0.096	-0.107	-	-0.107	0.002
	Gamma	11	0.325	0.264	0.277	0.334	0.263	0.102	0.040	0.022	-0.210	-0.098
	sh.=2	21	0.242	0.217	0.219	0.196	0.199	0.098	0.069	0.058	-0.012	-0.057
	sc.=2	51	0.172	0.163	0.161	0.139	0.133	0.098	0.087	0.082	0.051	-0.025
	Gamma	11	0.325	0.264	0.277	0.334	0.263	0.102	0.040	0.022	-0.210	-0.098
	sh.=2	21	0.242	0.217	0.219	0.196	0.199	0.098	0.069	0.058	-0.012	-0.057
0.9	sc.=0.5	51	0.172	0.163	0.161	0.139	0.133	0.098	0.087	0.082	0.051	-0.025
	Weibull	11	0.267	0.224	0.222	0.282	0.187	0.109	0.076	0.041	-0.167	-0.079
	sh.=2	21	0.214	0.196	0.189	0.173	0.138	0.117	0.103	0.082	-0.003	-0.043
	sc.=2	51	0.167	0.161	0.155	0.140	0.090	0.120	0.115	0.105	0.087	-0.019
	Weibull	11	0.267	0.224	0.222	0.282	0.187	0.109	0.076	0.041	-0.167	-0.079
	sh.=2	21	0.214	0.196	0.189	0.173	0.138	0.117	0.103	0.082	-0.003	-0.043
	sc.=0.5	51	0.167	0.161	0.155	0.140	0.090	0.120	0.115	0.105	0.087	-0.019

Table 4.10: Comparison between the confidence interval (Conf. Int.) and the Bayesian credible interval (Cred. Int.) in terms of coverage (Cov.), with the nominal level fixed at 0.95, and average width. Different combinations of σ^2 and n at different quantiles are reported.

		$p = 0.1$				$p = 0.25$			
		Conf. Int.		Cred. Int.		Conf. Int.		Cred. Int.	
σ^2	n	Cov.	Width	Cov.	Width	Cov.	Width	Cov.	Width
0.25	11	0.951	0.435	0.952	0.432	0.951	0.501	0.950	0.493
	21	0.949	0.310	0.948	0.309	0.949	0.350	0.952	0.349
	51	0.951	0.197	0.950	0.202	0.952	0.220	0.955	0.221
0.5	11	0.949	0.469	0.950	0.458	0.952	0.617	0.947	0.601
	21	0.951	0.335	0.952	0.334	0.950	0.431	0.950	0.428
	51	0.949	0.213	0.949	0.222	0.954	0.271	0.954	0.271
1	11	0.952	0.464	0.946	0.456	0.951	0.728	0.945	0.704
	21	0.950	0.329	0.949	0.331	0.951	0.504	0.951	0.500
	51	0.950	0.208	0.951	0.216	0.948	0.316	0.949	0.316
2	11	0.953	0.418	0.940	0.421	0.951	0.843	0.934	0.814
	21	0.954	0.283	0.954	0.284	0.953	0.559	0.949	0.554
	51	0.949	0.176	0.947	0.177	0.952	0.342	0.952	0.342
		$p = 0.50$				$p = 0.75$			
0.25	11	0.954	0.677	0.951	0.663	0.954	1.188	0.952	1.135
	21	0.950	0.457	0.948	0.455	0.949	0.749	0.948	0.744
	51	0.952	0.281	0.952	0.362	0.953	0.447	0.953	0.446
0.5	11	0.948	0.985	0.942	0.944	0.951	2.087	0.948	1.895
	21	0.948	0.655	0.946	0.650	0.951	1.262	0.951	1.241
	51	0.953	0.400	0.954	0.404	0.951	0.736	0.951	0.734
1	11	0.954	1.487	0.944	1.370	0.949	4.185	0.942	3.406
	21	0.949	0.953	0.946	0.938	0.951	2.324	0.950	2.237
	51	0.953	0.573	0.953	0.572	0.952	1.298	0.951	1.292
2	11	0.949	2.385	0.930	2.047	0.947	10.188	0.929	6.653
	21	0.949	1.429	0.943	1.383	0.949	4.866	0.944	4.454
	51	0.952	0.829	0.951	0.825	0.950	2.533	0.950	2.506
		$p = 0.90$				$p = 0.95$			
0.25	11	0.950	2.228	0.948	2.071	0.955	3.340	0.955	3.049
	21	0.949	1.316	0.949	1.301	0.949	1.859	0.949	1.833
	51	0.948	0.753	0.948	0.756	0.949	1.044	0.949	1.043
0.5	11	0.950	4.800	0.948	4.068	0.951	8.141	0.949	6.590
	21	0.949	2.578	0.948	2.508	0.947	4.010	0.947	3.878
	51	0.950	1.426	0.949	1.421	0.950	2.132	0.950	2.124
1	11	0.951	12.585	0.947	8.771	0.953	25.564	0.950	16.039
	21	0.949	5.894	0.948	5.527	0.946	10.542	0.946	9.723
	51	0.950	3.041	0.949	3.018	0.951	5.154	0.952	5.107
2	11	0.950	48.421	0.937	22.141	0.946	135.003	0.939	47.351
	21	0.952	17.162	0.950	14.683	0.953	37.937	0.950	31.048
	51	0.949	7.913	0.949	7.772	0.955	15.739	0.954	15.395

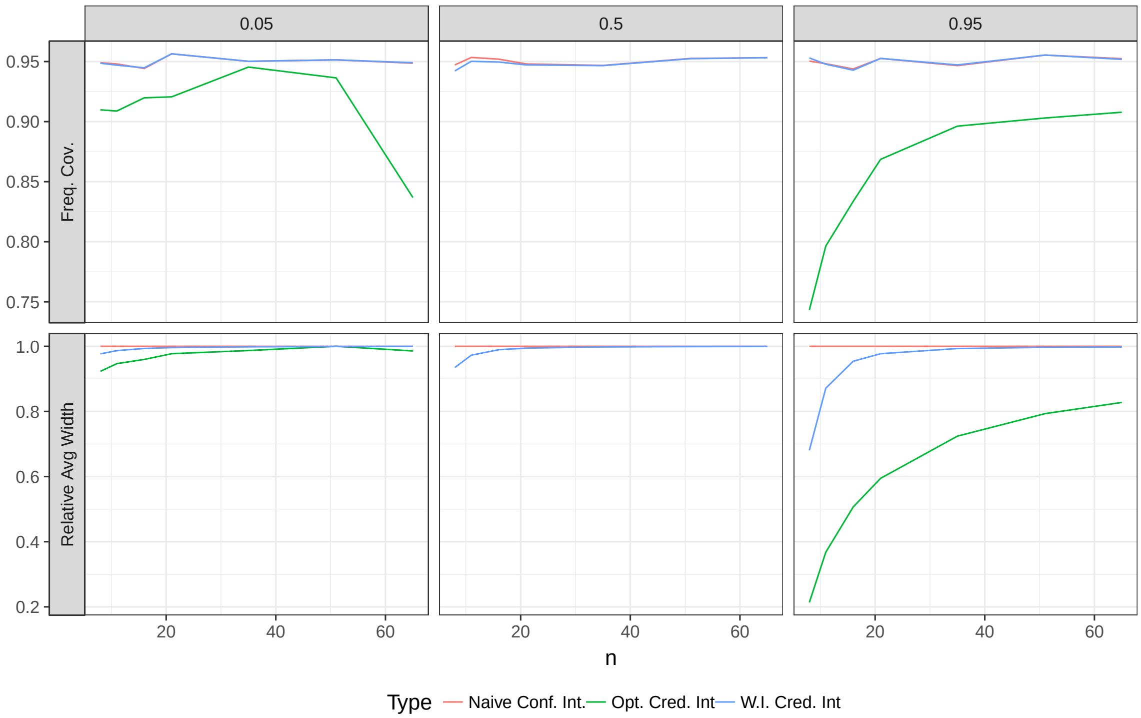

Figure 4.4: Average width of the Bayesian credible interval compared to the frequentist confidence interval. The trend with respect to the sample size n is reported at different quantiles p.

than the naive confidence intervals, indicating an important gain in precision. The results reported in this section are about two sided intervals; however simulations have been carried out for testing the behaviour of the one-sided intervals too and similar findings are obtained.

4.3 Quantile regression model

The assessment of the frequentist properties of the Bayes estimators used to estimate the log-normal quantile conditioned to a covariate point $\tilde{\mathbf{x}}$ is performed through a Monte Carlo study. The simulation scheme adopted is the same of the paper by Fabrizi and Trivisano (2016) and random samples from the following model are generated:

$$y_i|\tilde{x}_{1i} \sim \log\mathcal{N}\left(\beta_0 + \beta_1\tilde{x}_{1i}, \sigma^2\right),\ i = 1, ..., n; \tag{4.4}$$

where the coefficients vector is fixed as $(\beta_0, \beta_1) = (1, 1)$ and the covariate $\tilde{x}_{1i}$ is fixed equal to a value generated from a uniform distribution between 0 and 1. Different combinations of sample sizes and variances in the log-scale are considered, with $n = (11, 21, 51)$ and $\sigma^2 = (0.1, 0.25, 0.5, 1)$. To evaluate the estimators also outside the covariate range $[0, 1]$, the following points in the covariate space are considered: $\tilde{\mathbf{x}} = (1, x_0)$, with $x_0 = (0, 0.1, ..., 1.1, 1.2)$. The simulation study is based on $B = 10,000$ replicates.

The Bayes estimators are compared to the conditioned estimator $\hat{\theta}^*_p(\tilde{\mathbf{x}})$ defined in the (3.71), to the naive estimator:

$$\hat{\theta}_p(\tilde{\mathbf{x}}) = \exp\left\{\tilde{\mathbf{x}}^T\hat{\boldsymbol{\beta}} + \Phi^{-1}(p)\hat{\sigma}\right\}, \tag{4.5}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$ are the estimates obtained through the OLS procedure; and, finally, to the output of a particular case of the Box-Cox quantile regression $\hat{\theta}^{BO}_p(\tilde{\mathbf{x}})$. The latter estimation method consists in the back-transformation of the estimates obtained with the classical non-parametric quantile regression (Koenker, 2005) applied on the log-transformed data.
The Bayes estimators evaluated in the simulation when $p \neq 0.5$ are: the posterior expectation with optimal hyperparameters chosen numerically and average leverage $\bar{h}$, $\hat{\theta}^{QBn}(\tilde{\mathbf{x}})$, and the one with approximately optimal prior $\hat{\theta}^{QBa}(\tilde{\mathbf{x}})$. On the other hand, in the median case only, the Bayes estimator under relative quadratic loss with weakly informative prior is taken into account.
The outputs of the simulation study are graphical and illustrate the behaviour of the estimators relative RMSE (the conditioned estimator is considered as the reference value) and the relative bias with respect to different values of x_0.
In figures 4.5, 4.6 and 4.7 the results for quantiles 0.10, 0.50 and 0.90 are reported; whereas in figures D.1, D.2 and D.3 in the appendix, outputs which are related to other quantiles are showed.
As expected, in general, with low values of σ^2 and high sample size n the estimators tend to behave similarly. Moreover, the non-parametric procedure that produces the estimate $\hat{\theta}^{BO}_p(\tilde{\mathbf{x}})$ is always the worst estimator in terms of RMSE, especially in the boundaries of the covariate range. This estimator is positively biased.
Another estimator that is always positively biased is the naive one $\tilde{\theta}_p(\tilde{\mathbf{x}})$. However it results to be a competitive estimator, even if it is largely outperformed by the Bayes estimators in extreme scenarios.
The two versions of the Bayes estimator evaluated in this study produced good results: in general they behave very similarly, and they displayed better performances than the competitors, especially in critical situations. Moreover, the good results in term of RMSE of the Bayes estimators with high covariate values is attached with a considerable reduction of the bias. In particular, the negative bias that characterizes the Bayes estimator increase, producing an almost unbiased estimator when x_0 is above its sample range.
Finally, the performances of the credible intervals are assessed through a Monte Carlo study based on $B = 5,000$ simulations with the weakly informative prior setting. As a first step, a nominal coverage level of 0.90 is chosen and in figure 4.8 the frequentist coverages obtained in the simulation are reported. It is possible to observe that in critical scenarios ($n = 11$, $\sigma^2 > 0.25$ and x_0 at the extremes of the range at the same time), the intervals showed a lower coverage than the nominal one. On the other hand, no differences can be noticed with respect to the different quantiles considered.
Shifting the focus on the average width, in figure 4.9 its behaviour with respect to the covariate value x_0 in different combinations of n and σ^2 is showed. As expected, similar trends are noted in the different scenarios: intervals are narrower with low variance and high sample size, whereas they become wider as the quantile increase.

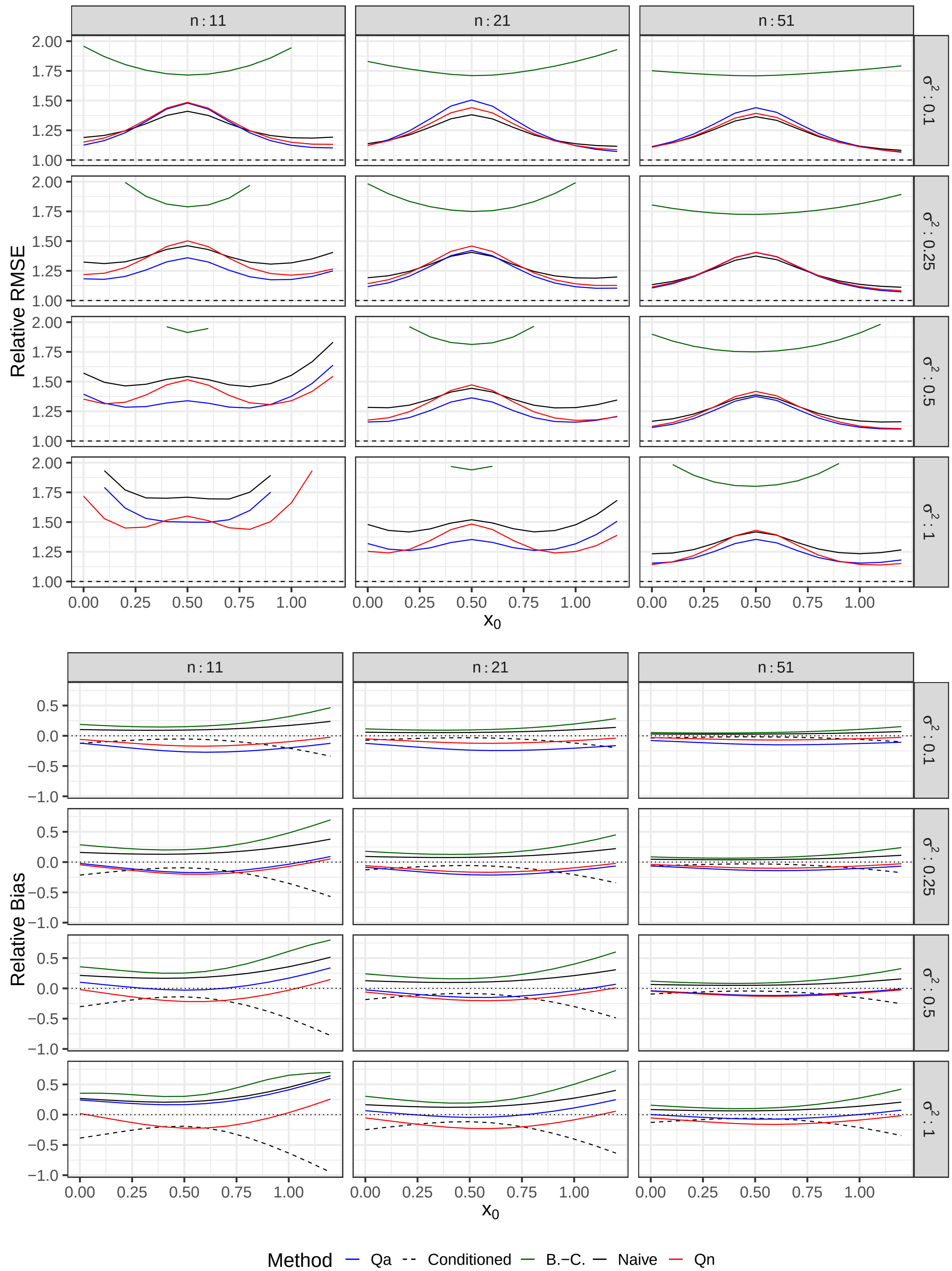

Figure 4.5: Relative RMSE and relative bias of various estimators of the target quantity $\theta_p(\tilde{\mathbf{x}})$, with $p = 0.10$.

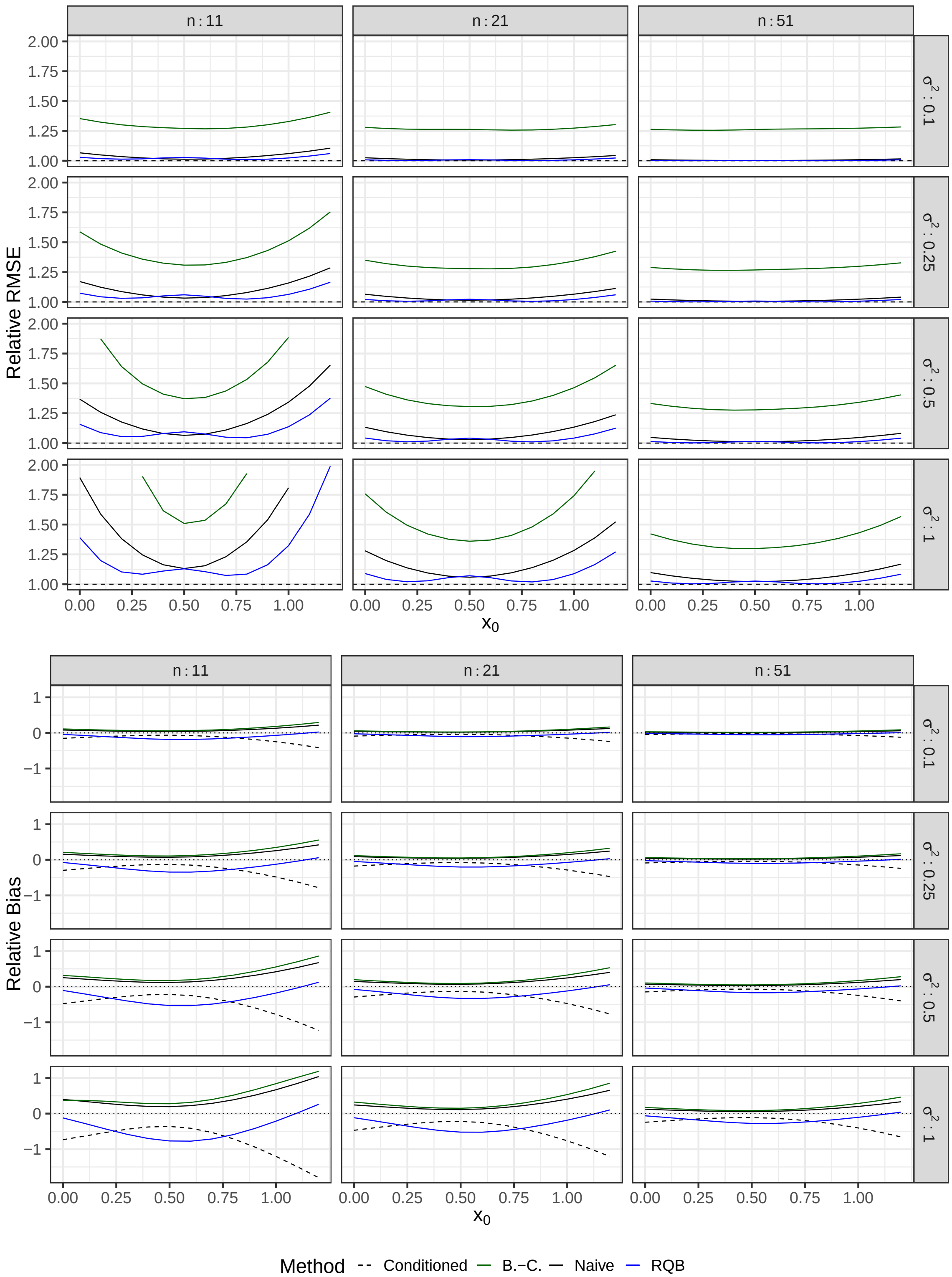

Figure 4.6: Relative RMSE and relative bias of various estimators of the target quantity $\theta_p(\tilde{\mathbf{x}})$, with $p = 0.50$.

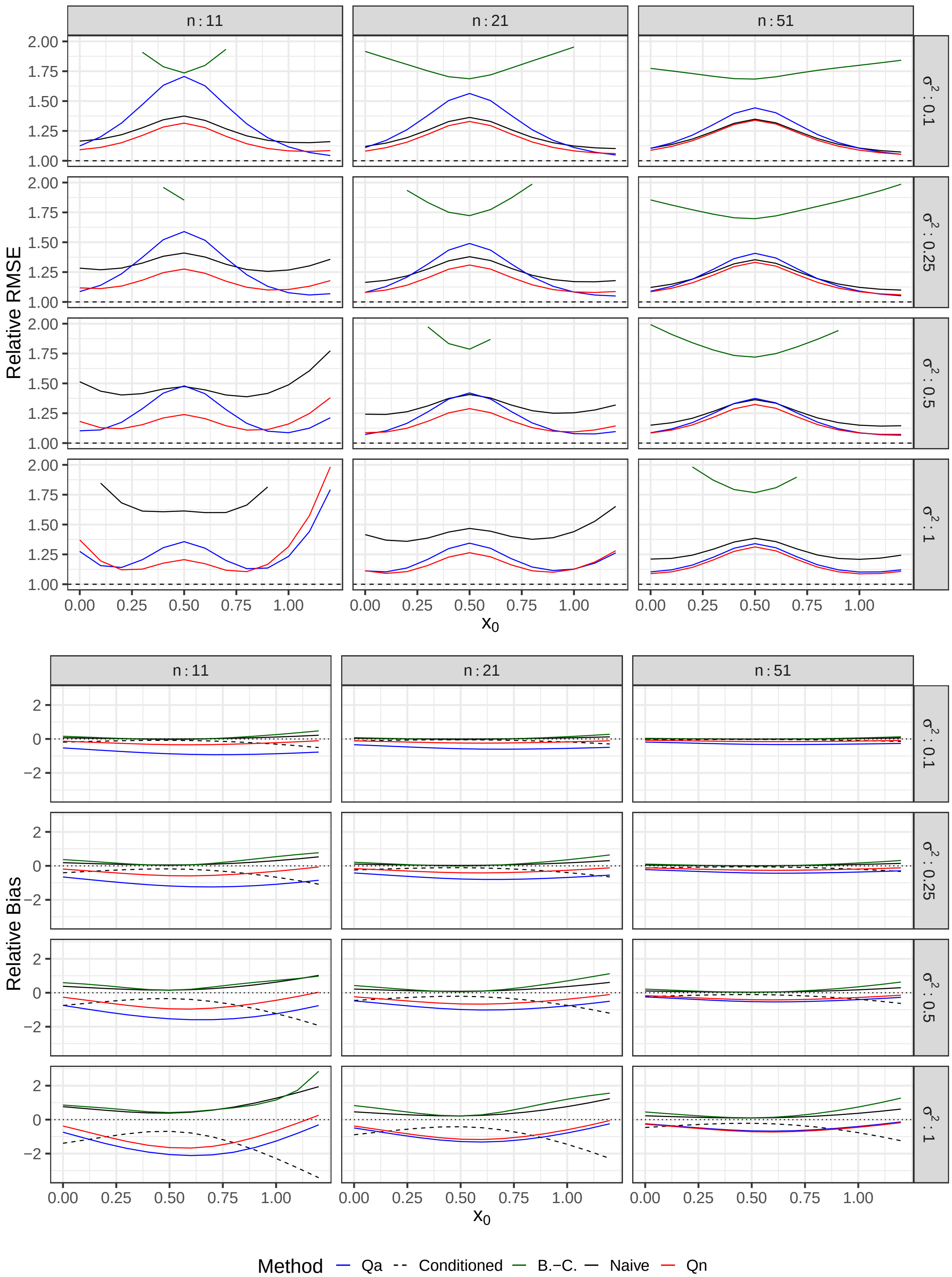

Figure 4.7: Relative RMSE and relative bias of various estimators of the target quantity $\theta_p(\tilde{\mathbf{x}})$, with $p = 0.90$.

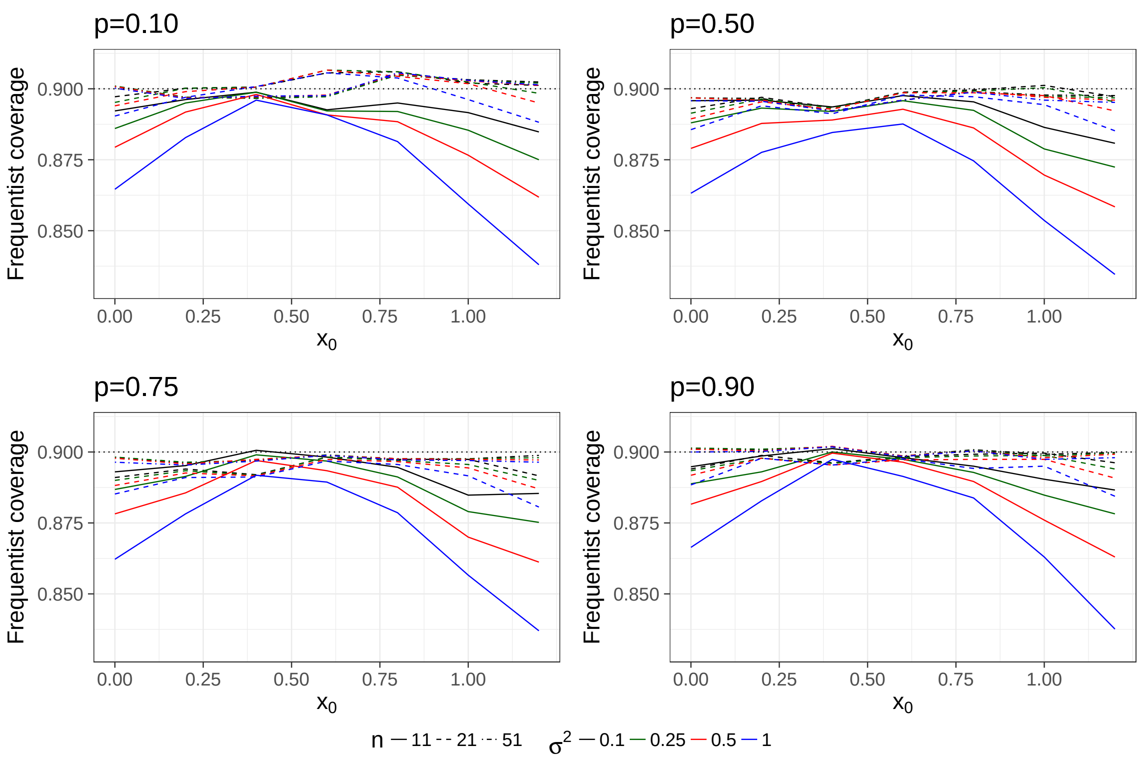

Figure 4.8: Frequentist coverages of the credible intervals. The nominal coverage is 0.90.

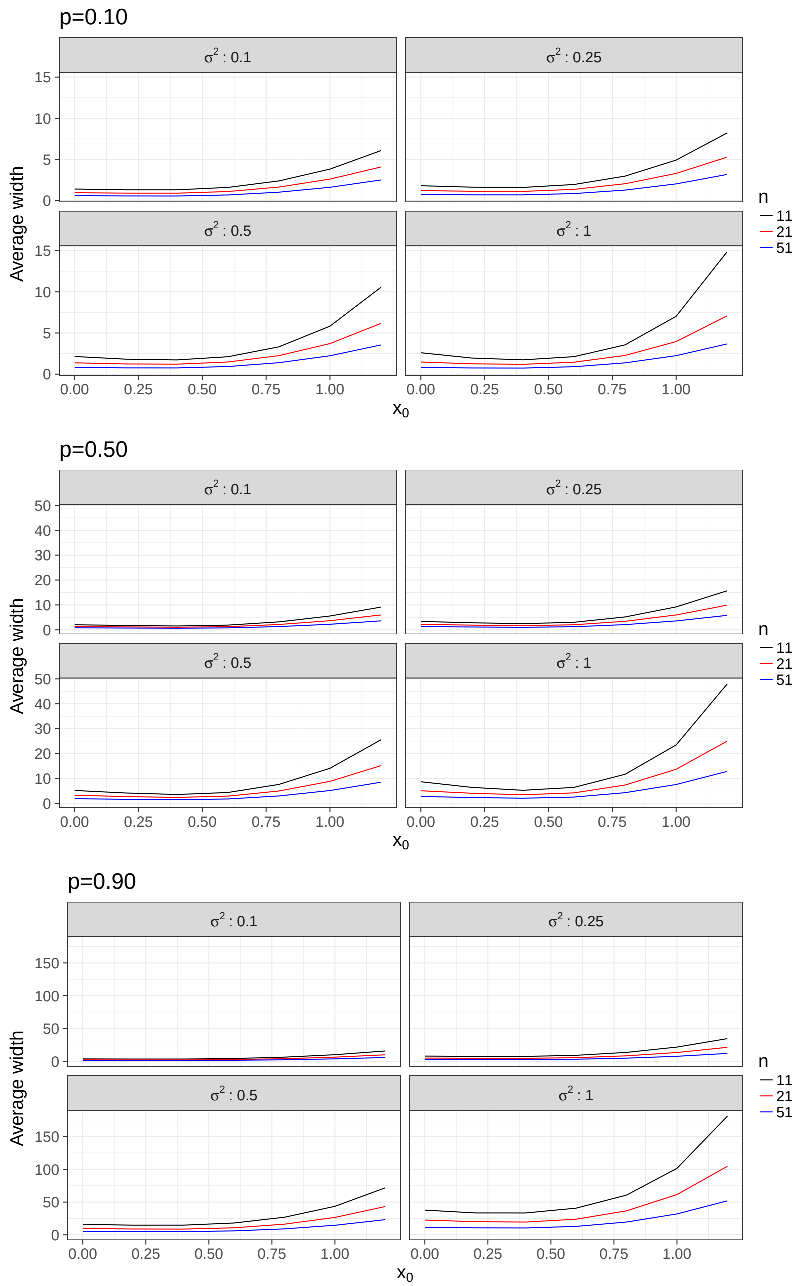

Figure 4.9: Average widths of the Bayesian credible intervals at different quantiles and combinations of σ^2 and n.

4.4 Examples

4.4.1 An application in environmental monitoring

In environmental monitoring, the estimation of a confidence interval around a percentile (usually in the right tail) is a common task. Besides, log-normality is often assumed in analysing pollutant concentration data. In fact, two different problems of this kind can be mentioned: the control for the presence of regulatory standard exceedences (that might be percentiles) and the estimation of the site background conditions with respect to a particular pollutant. In the second case, the estimated threshold is used to verify the compliance of subsequent samples and to assess the eventual contamination (USEPA, 2009). As already highlighted in section 3.1.2, in these applications, the one-sided lower and upper confidence intervals are often employed to respectively estimate the so called lower confidence limit (LCL) and upper confidence limit (UCL).
Since these procedures have a considerable impact both on the environmental protection and on the eventual corrective actions to afford, the improvements of statistical methods used in these procedures are particularly useful.
Besides, given that this kind of monitoring studies are often conducted with small sample sizes, the Bayesian approach for the construction of credible intervals around log-normal percentiles, whose frequentist performances are studied in section 4.2, might represent a substantial improvement in this context.
To illustrate the prospective improvements brought by the developed method, a popular example (USEPA, 2009; Millard, 2013) is faced. The original goal is to estimate the upper tolerance limit of the 95-th quantile, i.e. the background threshold that is represented by the value which includes the 95% of the distribution with the 95% confidence.
A small sample ($n = 8$) of chrysene concentrations (ppb) is obtained from two background wells:

$$19.7,\ 39.2,\ 7.8,\ 12.8,\ 10.2,\ 7.2,\ 16.1,\ 5.7.$$

All the most popular tests do not reject the hypothesis of log-normality.

Table 4.11: Point estimates of $\theta_{0.95}$ with different methods. The standard error estimates are reported too if available.

	$\hat{\theta}_{0.95}$	$\hat{Q}_{0.95}$	$\hat{\theta}_{0.95}^{QBw}$	$\hat{\theta}_{0.95}^{QBn}$	$\hat{\theta}_{0.95}^{QBa}$
Estimate	34.517	33.696	40.491	31.181	30.806
S.e.	-	8.056	26.862	8.257	9.581

As a first step, the 95-th percentile of the distribution is estimated assuming log-normality. The Bayes minimum MSE conditioned estimator $\hat{\theta}_{0.95}^{*}$ is not considered since the value of σ^2 is not known. Using the unbiased estimates of the mean and variance in the log-scale ($\hat{\mu} = 2.509$, $\hat{\sigma}^2 = 0.394$), the naive estimator $\hat{\theta}_{0.95}$ and the Longford's minimum MSE estimator $\hat{Q}_{0.95}$ are computed. Then, the Bayes estimator is evaluated under three different prior strategies: the weakly informative setting $\hat{\theta}_{0.95}^{QBw}$, the one with hyperparameters chosen by numerically minimizing the MSE $\left(\hat{\theta}_{0.95}^{QBn}\right)$ and the strategy that takes advantage of the analytic approximation of the work by Fabrizi and Trivisano (2012), $\hat{\theta}_{0.95}^{QBa}$. The weakly informative prior on the variance is given by $\sigma^2 \sim GIG(\lambda = 0, \delta = 0.01, \gamma = 1.06)$, as $n = 8$ and $\gamma_0 = 3/\sqrt{n}$); the numerically MSE-optimizing one is $\sigma^2 \sim GIG(\lambda = 0, \delta = 1, \gamma = 4.61)$ and the approximately optimal is $\sigma^2 \sim GIG(\lambda = -2, \delta = 0.01, \gamma = 1.06)$.

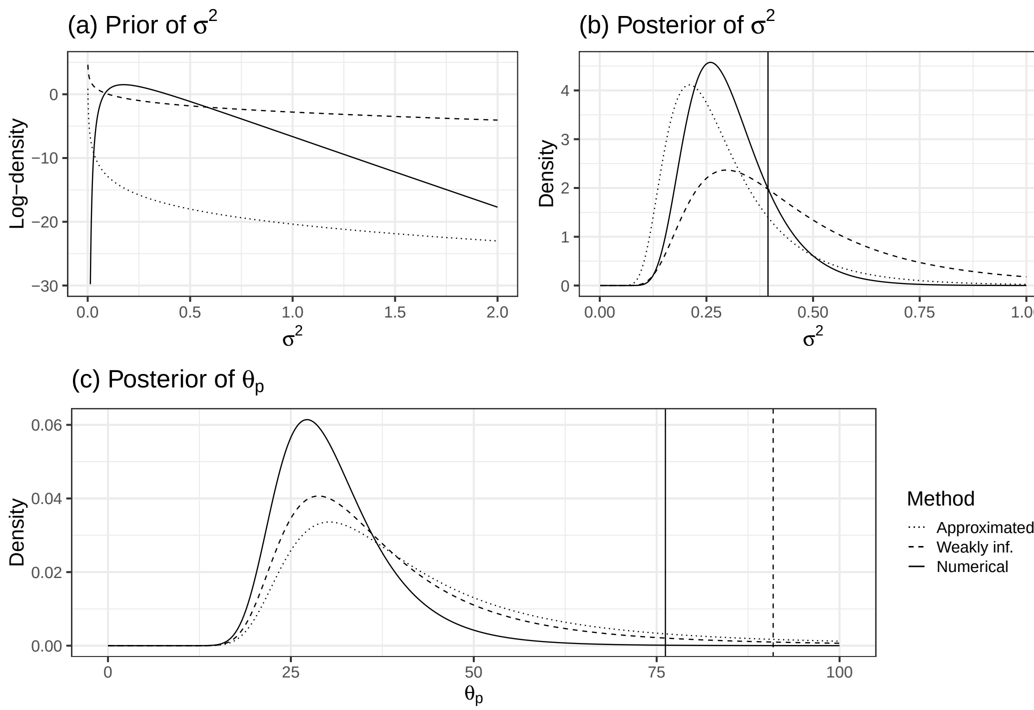

Figure 4.10: Comparison between the distributions obtained in the three different Bayesian approaches in the environmental monitoring example. *(a)* Log-density of the prior distributions of σ^2. *(b)* Density of the posterior distributions of σ^2, the vertical line represents the unbiased sample estimate $\hat{\sigma}^2$. *(c)* Posterior distributions of the target functional $\theta_{0.95}$. The solid vertical line reports the Bayesian estimate of the UCL, whereas the dashed one the frequentist estimate.

All the numerical results are reported in table 4.11, and in the case of the Bayes estimators the posterior standard deviation is reported too, in order to have an estimate of the estimator variability. According to the simulation results reported in section 4.1.4, the posterior variance is a reliable estimate of the estimator variance, even if a slight underestimation can be possible with small sample sizes and high quantiles. For Longford's $\hat{Q}_{0.95}$, the square root of the minimized MSE is reported as an estimate of the standard error, even if in extreme quantiles the author observed a severe underestimation of the true MSE.
As a natural consequence of the extreme inferential conditions of the problem, a moderate variability can be observed among the estimates. As expected, the estimators that are aimed

at minimizing the MSE assume the lowest values, especially $\hat{\theta}_{0.95}^{QBn}$ and $\hat{\theta}_{0.95}^{QBa}$, because of their negative bias, whereas the Bayes estimator having the weakly informative prior registers the higher value.
From the sub-figure 4.10.b, the underestimation of σ^2 produced by the prior settings of $\hat{\theta}_{0.95}^{QBn}$ and $\hat{\theta}_{0.95}^{QBa}$ is evident; whereas, in the weakly informative setting, the posterior $p(\sigma^2|\mathbf{x})$ is flatter and centred around $\hat{\sigma}^2$.
The heavy tail of the weakly informative prior distribution is inherited by the posterior of the target functional $p(\theta_p|\mathbf{x})$, as it is possible to deduce from figure 4.10 and from the posterior standard deviation reported in table 4.11. Conversely, the other two strategies produced peaked posteriors that are not suitable to produce interval estimates, as confirmed by the Monte Carlo study.
Considering the posterior quantiles of $p(\theta_p|\mathbf{x})$ in a weakly informative prior setting, the Bayesian credible interval can be provided. As studied in section 4.2, the constructed intervals possess an average width consistently lower than the frequentist intervals based on the normal theory, particularly with small n and extreme quantiles, maintaining the nominal coverage level. This characteristic holds also for one sided intervals, that are required for the determination of the UCL.
Lastly, the standard method currently employed by the EPA leads to an estimate of the background threshold equal to 90.925, whereas the Bayesian credible interval produces the value of 76.195. This implies that the procedure currently used to determine the UCL might be excessively conservative.
The results about the Bayesian estimation of $\theta_{0.95}$ under the optimal prior and the weakly informative one can be easily obtained using the `LN_Quant` function of the package `BayesLN`, where the used data are already present. The code and the outputs are now reported.

```
library(BayesLN)
# Load the dataset included in the package
data("EPA09")
EPA09

## [1] 19.7 39.2  7.8 12.8 10.2  7.2 16.1  5.7

# Optimal prior setting
LN_Mean(x = EPA09, x_transf = FALSE,
method = "optimal", CI = FALSE)

## ------------------
## Lognormal mean estimation with optimal prior
## ------------------
## $Prior_Parameters
## lambda  delta  gamma
## -3.800  0.010  2.031
##
```

```
## $Posterior_Parameters
## lambda  alpha  delta   beta     mu
## -7.300  7.000  0.587  4.000  2.509
##
## $LogN_Par_Post
##              Mean         Var    p=0.05    p=0.50    p=0.95
## xi     2.5085773 0.025427261 2.2479884 2.5085773 2.7691663
## sigma2 0.2034181 0.006442247 0.1089936 0.1867561 0.3538392
##
## $Post_Estimates
##           Mean     S.d.
## [1,] 13.79142 2.352631
```

```
# Weakly informative prior and UCL
LN_Mean(x = EPA09, x_transf = FALSE,
method = "weak_inf", alpha_CI = 0.05,
type_CI = "UCL")
```

```
## ------------------
## Lognormal mean estimation with weak_inf prior
## ------------------
## $Prior_Parameters
## lambda  delta  gamma
##  0.000  0.010  2.031
##
## $Posterior_Parameters
## lambda  alpha  delta   beta     mu
## -3.500  7.000  0.587  4.000  2.509
##
## $LogN_Par_Post
##              Mean        Var    p=0.05    p=0.50    p=0.95
## xi     2.5085773 0.04906591 2.1479301 2.5085773 2.8692245
## sigma2 0.3925273 0.03930270 0.1745759 0.3456645 0.7687825
##
## $Post_Estimates
##           Mean     S.d.
## [1,] 15.42667 4.241931
##
## $Interval
## Lower limit Upper limit
##      0.0000     22.9147
```

4.4.2 Application in occupational health

Also in public health analyses, it is common to estimate extreme quantiles in the right tail of a skewed distribution from small samples (Bullock and Ignacio, 2006; Gibbons et al., 2009; Krishnamoorthy et al., 2011).
To show a possible use of the developed methods in the occupational health research, a small dataset from the appendix IV of the book by Bullock and Ignacio (2006) is considered. It consists of $n = 10$ monitoring observations of exposure (ppm) of the coil feed operator and helper to Methyl Isobutyl Ketone (MIBK) during cleanup:

$$23,\ 42,\ 86,\ 62,\ 34,\ 107,\ 29,\ 65,\ 54,\ 55.$$

Likewise the previous example, the original inferential goal is the estimation of the 95-th percentile ($\theta_{0.95}$) of the underlying distribution and the associated upper tolerance limit. Then, this value needs to be compared to the MIBK short term exposure limit fixed at 75 ppm. The weakly informative prior on the variance is given by $\sigma^2 \sim GIG(\lambda = 0, \delta = 0.01, \gamma = 0.93)$, the one for $\hat{\theta}_{0.95}^{QBn}$ is given by $\sigma^2 \sim GIG(\lambda = 0, \delta = 1, \gamma = 6.24)$ and, finally, the prior for $\hat{\theta}_{0.95}^{QBa}$ is $\sigma^2 \sim GIG(\lambda = -2, \delta = 0.01, \gamma = 0.93)$: the prior settings are rather similar to the ones of the former application.
The different choices for the prior on σ^2 are reflected into different posteriors $p(\sigma^2|\mathbf{x})$: if a weakly informative choice is taken, the posterior would be more diffuse, thus giving more weight to large values of the estimates, while the two aimed at minimizing the MSE are more peaked around their mean and light-tail. These differences offer a clue to read the dissimilarities between estimators compared in table 4.12.

Table 4.12: Estimates of $\theta_{0.95}$ with different methods: naive ($\hat{\theta}_{0.95}$), Longford ($\hat{Q}_{0.95}$), the Bayes estimators under weakly informative prior ($\hat{\theta}_{0.95}^{QBw}$), MSE-optimizing prior ($\hat{\theta}_{0.95}^{QBn}$) and the approximately optimal prior ($\hat{\theta}_{0.95}^{QBa}$).

	$\hat{\theta}_{0.95}$	$\hat{Q}_{0.95}$	$\hat{\theta}_{0.95}^{QBw}$	$\hat{\theta}_{0.95}^{QBn}$	$\hat{\theta}_{0.95}^{QBa}$
Estimate	111.074	110.934	123.048	105.873	102.335
S.e.	-	17.771	44.554	19.626	22.514

For the estimation of the upper confidence limit, the Bayesian proposal, based on the weakly informative prior specification, is compared to the method currently employed in literature. The average shorter length of the intervals based on the proposed method leads to a smaller and more powerful estimate for the UCL (195.81 against 204.19).

4.4.3 Application in lifetime analysis

The log-normality assumption is also popular in the analysis of survival data: in this case the time of the event occurrence (e.g. failure, death, break) is the usual response. One of the goals of these kind of analysis is the estimation of low quantiles together with the

corresponding lower confidence limit, in order to have an idea of reliability. Another quantity of interest is the median, since the median duration is considered to be a performance indicator.
A popular example in this field (Lawless, 2003; Hahn and Meeker, 2011) is based on data from an endurance test on deep-groove ball bearings. It is a sample of size $n = 23$ and the observed result is the number of revolutions in millions:

17.88, 28.92, 33.00, 41.52, 42.12, 45.60, 48.40, 51.84, 51.96, 54.12, 55.56, 67.80,
68.64, 68.64, 68.88, 84.12, 93.12, 98.64, 105.12, 105.84, 127.92, 128.04, 173.40.

The log-normality assumption is not rejected according to most popular tests.
The unbiased estimates of the log-normal parameters are $\hat{\mu} = 4.150$ and $\hat{\sigma}^2 = 0.284$ and the same estimators reported in the previous examples are considered for the lower quantiles. On the other hand, in the median case, the Bayes estimator with the hyperparameters selected through the approximation cannot be used. Therefore, the Bayes estimator obtained under relative quadratic loss with weakly informative prior is considered.

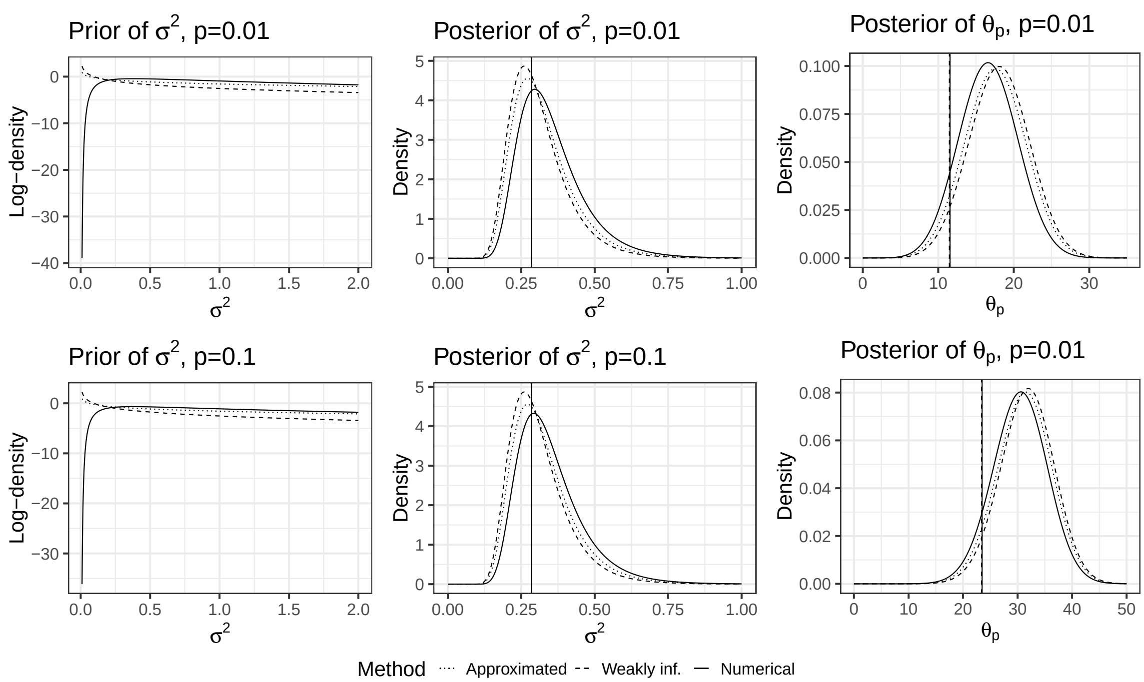

Figure 4.11: Comparison between the distributions obtained in the three different Bayesian approaches adopted for the example about lifetime data. For each quantile, the following plots are reported: log-density of the prior distributions of σ^2; density of the posterior distributions of σ^2, with the vertical line which represents the unbiased estimate $\hat{\sigma}^2$; posterior distributions of the target functional θ_p. In the last plot, the solid vertical line reports the Bayesian estimate of the LCL, whereas the dashed one the frequentist estimate.

As expected from the fact that the log-normal left tail is finite, the differences among the

estimates are only slight (table 4.13). In this framework, the naive estimator and the Bayes estimator under quadratic loss and weakly informative prior register the higher values too, whereas the estimators with smaller MSE assume lower values. Moreover, the underestimation of the $\hat{Q}_p$ standard error is clear also for the lower quantiles.
Looking at the results and at the distributions reported in figure 4.11, the likeness among different hyperparameters selection procedures is evident. However, the numerical optimization rule for the prior choice, which in this case is based on the minimization of the MSE with respect to δ, produces a GIG prior that is different from the others. In particular, by looking at the log-density plot, almost no probability is given to the 0 value.
As far as the median estimation concerns, some interesting cues can be deduced from figure 4.12. It can be noted that the Bayes estimator under quadratic loss (in blue) is slightly higher than the mode, causing a positive bias which makes the estimator to be not optimal, whereas the Bayes estimator under relative quadratic loss is lower than the peak.

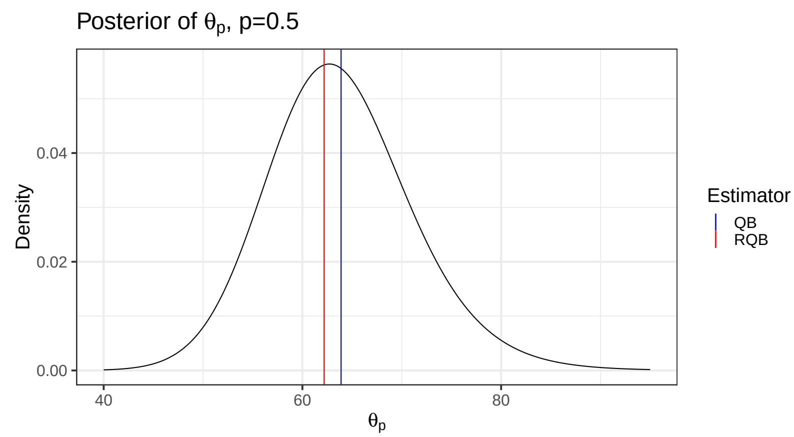

Figure 4.12: Posterior distribution of the target functional $\theta_{0.5}$. The solid vertical lines represent $\hat{\theta}_{0.5}^{QBw}$ (blue) and $\hat{\theta}_{0.5}^{RQBw}$ (red).

	$\hat{\theta}_p$		$\hat{Q}_p$		$\hat{\theta}_p^{QBw}$		$\hat{\theta}_p^{QBa}$		$\hat{\theta}_p^{QBn}$	
p	Est.	S.e.	Est.	S.e.	Est.	S.e.	Est.	S.e.	Est.	S.e.
0.01	18.347	-	17.486	2.174	18.065	3.948	17.706	3.955	16.660	3.866
0.1	32.034	-	30.950	3.626	31.708	4.907	31.357	4.949	30.416	4.983
	$\hat{\theta}_p$		$\hat{Q}_p$		$\hat{\theta}_p^{QBw}$		$\hat{\theta}_p^{RQBw}$			
0.5	63.458	-	62.267	7.036	63.889	7.469	62.180	7.469		

Table 4.13: Estimates of θ_p with different methods at $p \in \{0.01, 0.1, 0.5\}$. The estimates of the standard errors are reported.

The LCLs and the intervals with the nominal coverage level of 0.95 are reported in table 4.14

and they confirm that the Bayesian credible intervals are narrower than the intervals based on the exponentiation of the intervals derived in the normal theory. However, as already noted in the simulations, with the increase of the sample size and the quantiles below or equal to the median, the differences between the approaches are slighter and they almost coincide in this case.

	Confidence interval	Bayesian C.I.
$\text{LCL}_{0.01}$	11.476	11.552
$\text{LCL}_{0.1}$	23.416	23.486
C.I. $\theta_{0.5}$	[50.386;79.922]	[50.417;79.873]

Table 4.14: Confidence intervals and credible intervals (95%): LCL for the quantiles 0.01 and 0.1, two sided for the median.

4.4.4 Application in hydrology

An important branch of hydrology is flood frequency analysis (Hamed and Rao, 1999), whose principal aim is the estimation of the magnitude of future extreme rivers flow.
Among the adopted approaches to this inferential problem, a widespread procedure consists of analysing the observed annual maximum flow time series of the studied river. Then, the yearly maximum flow with a fixed return period T is estimated. This analysis is carried out assuming that the observed maxima are independent and identically distributed according to a positively skewed distribution.
To link this problem to the topic of this work, it is worth to stress that an event having a fixed return period T is equivalent to the $\left(1 - \frac{1}{T}\right)$-th quantile of the distribution. Moreover, even if the most recent guidelines for flood frequency analysis (England Jr et al., 2018) consider the log-Pearson type III distribution as automatic assumption for this kind of data, the log-normal distribution is widely used in this research area (Stedinger, 1980; Karim and Chowdhury, 1995; Vogel and Wilson, 1996; Strupczewski et al., 2001).
In order to exploit the efficiency of the proposed estimation method in small samples inference, a river recently monitored is chosen for this example. The series of the annual maximum peak of Peeptoad Brook (1994-2016), a river in the Rohde Island State (USA), was obtained from the *United States Geological Survey* (USGS), through the *National Water Information System*.
The most popular tests do not reject the log-normality assumptions for this set of data.
The outputs about a set of quantiles connected to frequently estimated return periods are reported in table 4.15. Considering the estimators behaviours, similar results to the ones of the application in environmental monitoring in section 4.4.1 are found. In particular, the higher estimates are registered by the Bayes estimator under relative quadratic loss, whereas the Bayes estimators aimed at minimizing the MSE assume the lowest.
Thereafter, the stationarity assumption of the observed flow maxima series can be relaxed by exploring the presence of relations with a set of covariates (Villarini et al., 2009; Vogel et al.,

Table 4.15: Estimates of θ_p with different methods at $p \in \{0.80, 0.90, 0.98, 0.99\}$. The estimates of the standard errors are reported.

		$\hat{\theta}_p$		$\hat{Q}_p$		$\hat{\theta}_p^{QBw}$	
T	p	Est.	S.e.	Est.	S.e.	Est.	S.e.
5	0.80	241.211	-	239.220	33.991	247.059	34.215
10	0.90	300.846	-	299.756	44.610	311.653	51.660
50	0.98	443.361	-	445.243	72.482	470.256	107.310
100	0.99	508.406	-	511.940	86.206	544.395	138.131

		$\hat{\theta}_p^{QBa}$		$\hat{\theta}_p^{QBn}$	
T	p	Est.	S.e.	Est.	S.e.
5	0.80	236.724	28.551	235.580	26.366
10	0.90	291.984	41.099	292.105	37.197
50	0.98	422.995	79.101	424.818	68.165
100	0.99	482.490	99.277	484.468	84.053

2011; Prosdocimi et al., 2014). In this framework, the inferential problem consists in the estimation of extreme quantiles conditioned to the values of a set of covariates. In particular, following the idea of the latter references, the behaviour of the annual maximum flood with respect to the year is explored. In this framework, the location of the log-transformation of the annual peak flood y_i is modelled using the year t_i as explanatory variable:

$$\log[y_i] = \beta_0 + \beta_1(t_i - \bar{t}), \tag{4.6}$$

where $\bar{t}$ is the average year.
The Bayes estimators are compared to the naive procedure (i.e. OLS estimation of the vector $\boldsymbol{\beta}$ and exponentiation of the normal quantile). The usual prior proposals for σ^2 are considered: the weakly informative $\left(\hat{\theta}_p^{QBw}(\tilde{\mathbf{x}})\right)$, the one based on the approximation by Fabrizi and Trivisano (2016) $\left(\hat{\theta}_p^{QBa}(\tilde{\mathbf{x}})\right)$ and the numerical optimization choice $\left(\hat{\theta}_p^{QBn}(\tilde{\mathbf{x}})\right)$. Moreover, for the weakly informative setting, the 90% credible intervals are also reported as a credibility region. In fact, it is an useful quantity that is natural to obtain in the Bayesian context but that is not reported in the applications involving the naive estimation method. The estimates of the return periods 5,10,100 (i.e. quantiles 0.80, 0.90, 0.99) are computed for the observed times and the following 5 years.
As it is possible to see from the marginal posterior distributions of the regression coefficient β_1, reported in figure 4.13, $\hat{\theta}_p^{QBa}(\tilde{\mathbf{x}})$ has the most peaked distribution, whereas, as expected, $\hat{\theta}_p^{QBw}(\tilde{\mathbf{x}})$ has the heavier tails. In the middle, the three choices of γ obtained for the different quantiles through the numerical optimization produce very similar distributions.
The behaviour of the estimates at different years are shown in figure 4.14. The credibility intervals reported as shaded areas indicate the increasing uncertainty in the estimation of

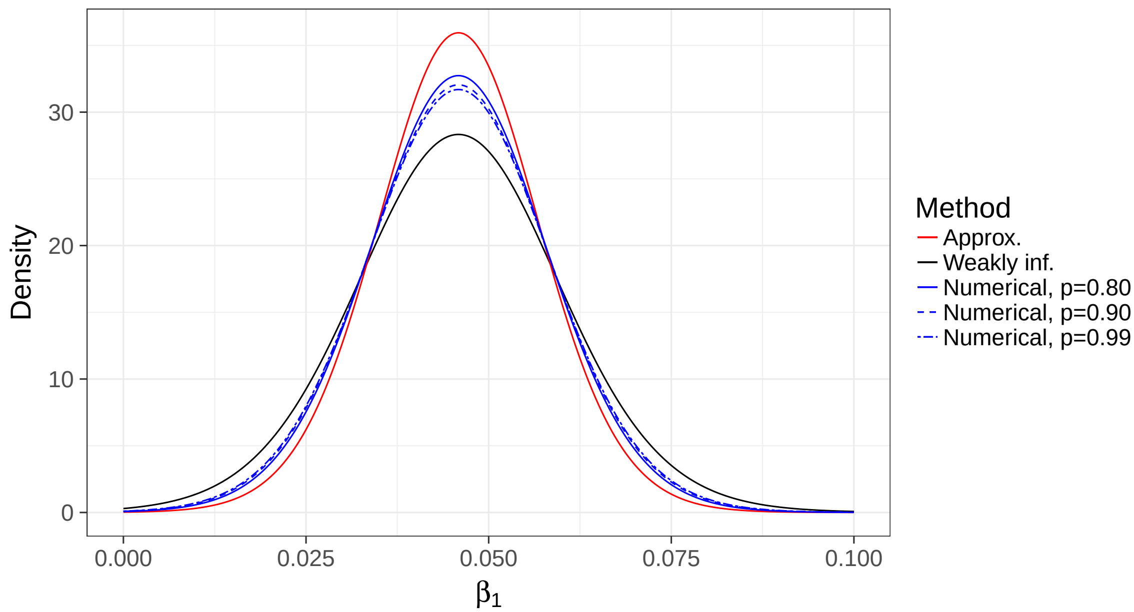

Figure 4.13: Posterior density of the regression coefficient β_1 under different prior distributions for the variance σ^2.

future quantiles. On the other hand, considering the point estimates, the observed behaviour is similar to the one proper of the unconditional quantile estimation: the higher values are obtained by the naive estimation and by $\hat{\theta}_p^{QBw}(\tilde{\mathbf{x}})$, whereas $\hat{\theta}_p^{QBa}(\tilde{\mathbf{x}})$ and $\hat{\theta}_p^{QBn}(\tilde{\mathbf{x}})$ are almost overlapped and produce the lowest estimates.

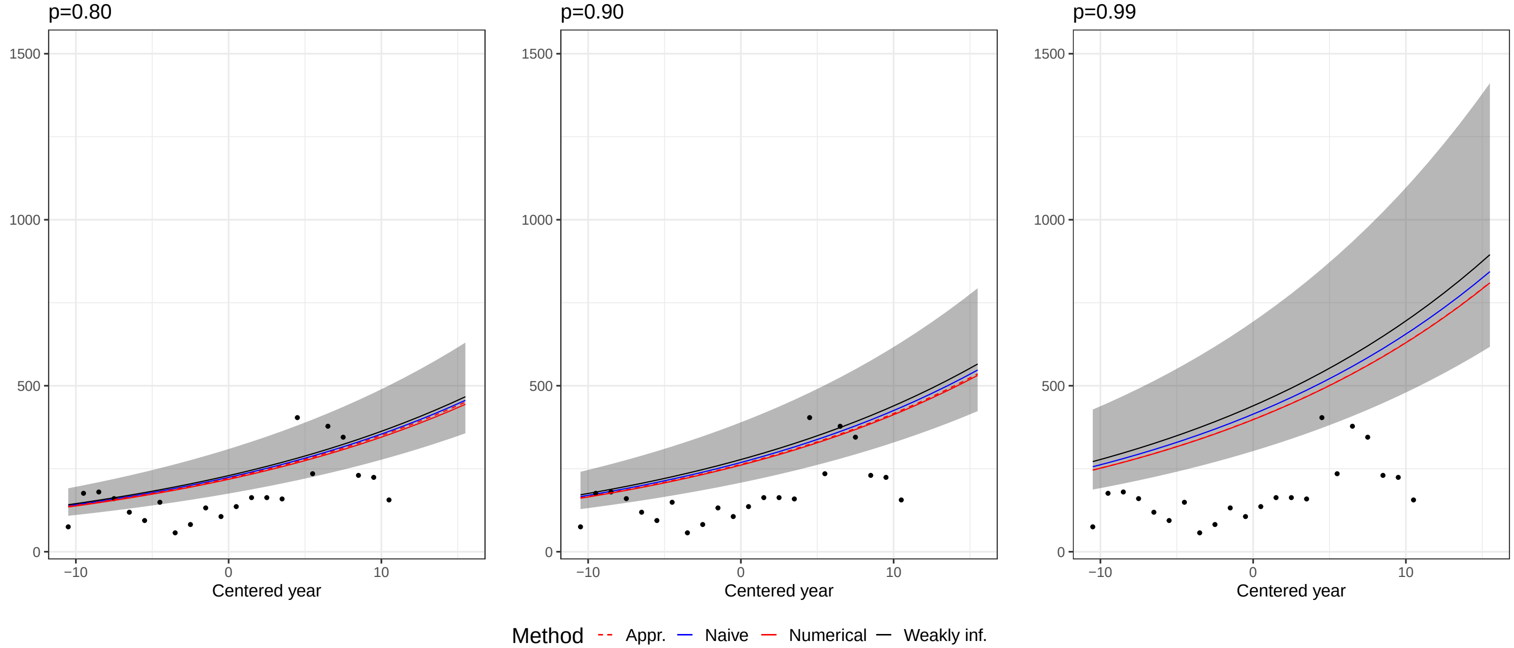

Figure 4.14: Estimated conditional quantiles of the annual peak flow with different methods and different probability levels. The shaded area represents the 90% credible interval for the weakly informative prior setting.

Part II

The log-normal linear mixed model

Chapter 5

Bayesian analysis of the log-normal linear mixed model

The log-normal linear mixed model is strictly linked to the well known normal model, which has received a lot of attention in the literature: many methodological aspects have been already deeply analysed both from the distributional viewpoint (Broemeling, 1985; Harville and Zimmermann, 1996) and from the prior specification issue (Gelman, 2006; Polson et al., 2012). The aim of this part of thesis is to focus on the log-normal linear mixed model and the estimation of quantities in the original data scale, such as the conditional expectation given the random effects or the marginalized one, deriving their posterior properties. In order to fulfil this task, the starting point is a review of the posterior analysis of a normal linear mixed model, then the original findings of this work will be stated.
In Bayesian inference, it is widely known that improper priors might be specified without loosing the properness of the posterior distribution. However, a mathematical check of the posterior is always requested because possible problems, like posterior improperness or posterior moments infiniteness, could be masked by the use of numerical techniques like Monte Carlo methods (Hobert and Casella, 1996), as already hinted in the introduction.
In the same spirit of the conclusions by Hobert and Casella (1996) about the unreliability of performing the Gibbs sampler when the posterior distribution is not proper, practitioners should be careful in computing moments from the posterior distributions. In fact, when the integral which defines the moment is not finite, the outcome from MCMC algorithms might appear to be reasonable even if it is completely meaningless. In the literature, warnings about this potential issue were highlighted at the beginning of the diffusion of Monte Carlo methods in Bayesian statistics (Geweke, 1989), then this topic attracted a lot of attention in the last years, since some priors where found to produce posteriors without finite moments in particular conditions (Fernandez and Steel, 2000; Sun and Speckman, 2005; Ghosh et al., 2018). Moreover, these eventualities might be even more unexpected since they may occur with proper priors too.
However, it must be stressed that, in order to apply powerful computational methods like MCMC, some theoretical properties about the problem of interest must always be checked.

Particularly, to compute an integral through Monte Carlo methods (e,g. what is carried out in Bayesian inference when the posterior mean is computed from the generated samples) it is requested that the integral considered is analytically finite.
The chapter is organized as follows. Firstly, the results concerning the simple one-way random effect model will be considered (section 5.1), then the general log-normal mixed model will be faced (section 5.2). A new prior specification strategy will be provided in section 5.3 and, finally, some details about the application of the developed methodologies in the small area estimation context will be given in section 5.4.

5.1 One-way ANOVA random effect model

The one-way ANOVA random effect model is very popular and diffused among applied scientists. Moreover, in many applications, it is common to consider the logarithmic transformation of the response variable. In this section, the unbalanced case will be considered, generalizing the model described in the introduction through the (1.9). The estimation of the normal one-way random effect model in the Bayesian framework has been studied in the works by Tiao and Tan (1965) and Hill (1965), in which the posterior distributions of the main parameters were deduced and, if necessary, approximated. Here, the main focus is the definition of quantities of interest in the log-normal context and the related posterior properties.
A sample of observations y_{ij} structured in $j = 1, ..., m$ groups with sample size n_j, with $\sum_j n_j = n$, is considered. In this context, the one-way random effect model is specified for the natural logarithm of the observed variable $w_{ij} = \log(y_{ij})$:

$$w_{ij} = \log(y_{ij}) = \mu + \nu_j + \varepsilon_{ij}, \tag{5.1}$$

assuming that the random effects ν_j are distributed as $\mathcal{N}\left(0, \tau^2\right)$ and are independent with respect to the unstructured error terms ε_{ij}, for which a $\mathcal{N}\left(0, \sigma^2\right)$ is assumed.
This model might be naturally specified as a three stages hierarchical Bayesian model with likelihood, priors of the parameters $(\mu, \boldsymbol{\nu}, \sigma^2)$ and a hyperprior for τ^2:

$$\begin{aligned} &\log(y_{ij})|\mu, \boldsymbol{\nu}, \sigma^2, \sim \mathcal{N}\left(\mu + \nu_j, \sigma^2\right); \\ &\nu_j|\mu, \sigma^2, \tau^2 \sim \mathcal{N}\left(0, \tau^2\right), \ \forall j; \qquad \left(\mu, \sigma^2\right) \sim p\left(\mu, \sigma^2\right); \\ &\tau^2 \sim p\left(\tau^2\right). \end{aligned} \tag{5.2}$$

Even if one of the main aims in applying this ANOVA model is the estimation of the variance components, often, for understanding and explanatory purposes, summary quantities in the original data scale are of interest. If the inferential goal is the estimation of an overall population mean, the vector of random effects is integrated out and the following results immediately follow:

$$\mathbb{E}\left[\log(y_{ij})|\mu, \sigma^2, \tau^2\right] = \mu; \quad \mathbb{V}\left[\log(y_{ij})|\mu, \sigma^2, \tau^2\right] = \sigma^2 + \tau^2. \tag{5.3}$$

Therefore, recalling the basic properties of the log-normal distribution, the marginal expectation of the observed random variable Y in its original scale is:

$$\theta_m = \mathbb{E}[y_{ij}|\mu, \sigma^2, \tau^2] = \exp\left\{\mu + \frac{\sigma^2 + \tau^2}{2}\right\}, \ \forall j, \ \forall i. \tag{5.4}$$

Alternatively, if a group-specific mean is required, it is possible to consider the following functional that is conditioned to the l-th group specific random effect ν_l:

$$\theta_c(\nu_l) = \mathbb{E}[y_{il}|\mu, \nu_l, \sigma^2] = \exp\left\{\mu + \nu_l + \frac{\sigma^2}{2}\right\}, \ \forall i. \tag{5.5}$$

From a Bayesian viewpoint, these two quantities might be estimated simply through the mean of their posterior distributions. However, as it will be illustrated later, the existence of the posterior mean and, consequently, of higher order moments, is not assured, even if the priors for the variance components are proper.
To fix some useful notation for sample quantities, the arithmetic mean of the l-th group units and the sum of squares within the groups are defined as:

$$\bar{w}_{.l} = \frac{\sum_{i=1}^{n_l} w_{il}}{n_l}, \qquad SSW = \sum_{l=1}^{m}\sum_{i=1}^{n_l}\left(w_{il} - \bar{w}_{.l}\right)^2. \tag{5.6}$$

The derivation of the l-th group specific likelihood function is immediate:

$$\begin{aligned} L(\nu_l, \sigma^2, \mu, \tau^2) &= \prod_{i=1}^{n_l} f(w_{il}|\mu, \sigma^2, \tau^2, \nu_l) \\ &\propto \left(\sigma^2\right)^{-\frac{n_l}{2}} \exp\left\{-\frac{n_l}{2\sigma^2}\left(\bar{w}_{.l} - \mu - \nu_l\right)^2\right\} \times \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_l}\left(w_{il} - \bar{w}_{.l}\right)^2\right\}; \end{aligned} \tag{5.7}$$

and the likelihood marginalized with respect to the random effects is:

$$\begin{aligned} L(\sigma^2, \mu, \tau^2) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{j=1}^{m} L(\nu_j, \sigma^2, \mu, \tau^2) \prod_{j=1}^{m} f(\nu_j|\mu, \sigma^2, \tau^2) \mathrm{d}\boldsymbol{\nu} \\ &\propto \left(\sigma^2\right)^{-\frac{n-m}{2}} \left[\prod_{j=1}^{m}\left(\sigma^2 + n_j\tau^2\right)^{-\frac{1}{2}}\right] \times \\ &\quad \times \exp\left\{-\frac{1}{2}\left(\sum_{j=1}^{m}\frac{n_j\left(\bar{w}_{.j} - \mu\right)^2}{(\sigma^2 + n_j\tau^2)} + SSW\right)\right\}. \end{aligned} \tag{5.8}$$

Setting a flat improper prior for the overall mean in the log-scale $p(\mu) \propto 1$ and two independent and generic priors for the variance components parameters:

$$p(\sigma^2, \tau^2) = p(\sigma^2)p(\tau^2), \tag{5.9}$$

the following posterior distributions, derived with standard computations, are useful for the subsequent results of this section:

$$\nu_l|\mu,\sigma^2,\tau^2,\mathbf{w} \sim \mathcal{N}\left(\bar{\nu}_l, V_{\nu_l}\right), \ l=1,...,m; \tag{5.10}$$

$$\mu|\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{w} \sim \mathcal{N}\left(\frac{\sum_{j=1}^m n_j\left(\bar{w}_{\cdot j}-\nu_j\right)}{n}, \frac{\sigma^2}{n}\right); \tag{5.11}$$

$$\mu|\sigma^2,\tau^2,\mathbf{w} \sim \mathcal{N}\left(\bar{\mu}, V_{\mu}\right); \tag{5.12}$$

$$\begin{aligned} p(\mu,\sigma^2,\tau^2|\mathbf{w}) \propto\ & p(\sigma^2)p(\tau^2)\left(\sigma^2\right)^{-\frac{n-m}{2}}\left[\prod_{j=1}^m\left(\sigma^2+n_j\tau^2\right)^{-\frac{1}{2}}\right]\exp\left\{-\frac{SSW}{2\sigma^2}\right\}\times \\ & \times\exp\left\{-\frac{1}{2}V_\mu^{-1}\left(\hat{\mu}-\mu\right)^2\right\}\exp\left\{-\frac{1}{2}\sum_{j=1}^m\frac{n_j\left(\bar{w}_{\cdot j}-\hat{\mu}\right)^2}{(\sigma^2+n_j\tau^2)}\right\}; \end{aligned} \tag{5.13}$$

$$\begin{aligned} p(\sigma^2,\tau^2|\mathbf{w}) \propto\ & p(\sigma^2)p(\tau^2)\left(\sigma^2\right)^{-\frac{n-m}{2}}V_\mu^{\frac{1}{2}}\left[\prod_{j=1}^m\left(\sigma^2+n_j\tau^2\right)^{-\frac{1}{2}}\right]\exp\left\{-\frac{SSW}{2\sigma^2}\right\}\times \\ & \times\exp\left\{-\frac{1}{2}\sum_{j=1}^m\frac{n_j\left(\bar{w}_{\cdot j}-\hat{\mu}\right)^2}{(\sigma^2+n_j\tau^2)}\right\}; \end{aligned} \tag{5.14}$$

where mean and variance of the full conditional of ν_l are:

$$\bar{\nu}_l = \frac{\frac{n_l}{\sigma^2}\left(\bar{w}_{\cdot l}-\mu\right)}{\frac{1}{\tau^2}+\frac{n_l}{\sigma^2}}, \qquad V_{\nu_l} = \frac{\tau^2\sigma^2}{\sigma^2+n_l\tau^2}; \tag{5.15}$$

and considering the parameters of the posterior of μ conditioned with respect to the variance components:

$$\bar{\mu} = \frac{\sum_{j=1}^m\frac{n_j\bar{w}_{\cdot j}}{(\sigma^2+n_j\tau^2)}}{\sum_{j=1}^m\frac{n_j}{(\sigma^2+n_j\tau^2)}}, \qquad V_\mu = \left(\sum_{j=1}^m\frac{n_j}{(\sigma^2+n_j\tau^2)}\right)^{-1}. \tag{5.16}$$

Performing a simple linear transformation of the (5.10) and the (5.11), the posterior distributions of the target functionals θ_m and $\theta_c(\nu_l)$ conditioned with respect to the variance components are:

$$\theta_m|\sigma^2,\tau^2,\mathbf{w} \sim \log\mathcal{N}\left(\bar{\mu}+\frac{\sigma^2+\tau^2}{2}, V_\mu\right), \tag{5.17}$$

$$\theta_c(\nu_l)|\boldsymbol{\nu},\sigma^2,\tau^2,\mathbf{w} \sim \log\mathcal{N}\left(\frac{\sum_{j=1}^m n_j\left(\bar{w}_{\cdot j}-\nu_j\right)}{n}+\nu_l+\frac{\sigma^2}{2}, \frac{\sigma^2}{n}\right). \tag{5.18}$$

Following the same reasoning of the first part of the thesis, in order to have a flexible prior for the variance component, two independent GIG priors (see section 2.1) are specified for σ^2 and τ^2:

$$p(\sigma^2) \sim GIG\left(\lambda_\sigma, \delta_\sigma, \gamma_\sigma\right), \quad p(\tau^2) \sim GIG\left(\lambda_\tau, \delta_\tau, \gamma_\tau\right). \tag{5.19}$$

The main results about the existence of the Bayes estimator under quadratic loss of the functionals θ_m (5.4) and $\theta_c(\nu_l)$ (5.5) are synthesized in the following theorem.

Theorem 5.1. *Considering the one way random effect model described in* (5.2) *with a flat improper prior on* μ *and the GIG priors* (5.19) *for the variance components, then:*

(i) the posterior moments of $\theta_c(\nu_l)$, $l = 1, ..., m$; *are define up to the order* $r > 0$ *if:*

$$\gamma_\sigma^2 > r + \frac{r^2}{n}; \tag{5.20}$$

(ii) the posterior moments of θ_m *are define up to the order* $r > 0$ *if:*

$$\begin{aligned} \gamma_\sigma^2 &> r + \frac{r^2}{n}, \\ \gamma_\tau^2 &> r + \frac{r^2}{m}. \end{aligned} \tag{5.21}$$

Proof. (i) It is necessary to find the existence conditions for the integral that defines the r-th moment:

$$\begin{aligned} \mathbb{E}\left[\theta_c(\nu_l)^r|\mathbf{w}\right] &= \mathbb{E}\left[\exp\left\{r\mu + r\nu_l + r\frac{\sigma^2}{2}\right\}|\mathbf{w}\right] \\ &= \int_\Theta \exp\left\{r\mu + r\nu_l + r\frac{\sigma^2}{2}\right\} p(\nu_l|\mu, \sigma^2, \tau^2, \mathbf{w}) \times \\ &\quad \times p(\mu|\sigma^2, \tau^2, \mathbf{w}) p(\sigma^2, \tau^2|\mathbf{w}) \mathrm{d}\boldsymbol{\theta}. \end{aligned} \tag{5.22}$$

The first step consists in solving the integral in ν_l. The integral is straightforward since it is the moment generating function of a normal distribution:

$$\int_{\mathbb{R}} \exp\left\{r\nu_l\right\} p_{\mathcal{N}}\left(\nu_l; \bar{\nu}_l, V_{\nu_l}\right) \mathrm{d}\nu_j = r\bar{\nu}_l + r^2\frac{V_{\nu_l}}{2} = r\frac{\frac{n_l}{\sigma^2}\bar{w}_{.l}}{\frac{1}{\tau^2} + \frac{n_l}{\sigma^2}} - r\frac{\frac{n_l}{\sigma^2}\mu}{\frac{1}{\tau^2} + \frac{n_l}{\sigma^2}} + r^2\frac{V_{\nu_l}}{2}. \tag{5.23}$$

Now, defining $rs_0\mu = r\mu - r\frac{\frac{n_l}{\sigma^2}\mu}{\frac{1}{\tau^2} + \frac{n_l}{\sigma^2}}$, where $s_0 = \frac{\sigma^2}{\sigma^2 + n_l\tau^2}$; it is possible to integrate out μ, having a normal MGF again:

$$\int_{\mathbb{R}} \exp\left\{rs_0\mu\right\} p_{\mathcal{N}}\left(\mu; \bar{\mu}, V_\mu\right) \mathrm{d}\mu = rs_0\bar{\mu} + r^2 s_0^2\frac{V_\mu}{2}. \tag{5.24}$$

Defining with $g(\sigma^2, \tau^2)$ a function that does not affect the finiteness of the integral, the (5.22) can now be written as:

$$\int_0^{+\infty}\int_0^{+\infty} g(\sigma^2,\tau^2)\exp\left\{-\frac{1}{2}\sigma^2\left(\gamma_\sigma^2 - r - r^2\left[\frac{V_{\nu_l}}{\sigma^2}+\frac{s_0^2 V_\mu}{\sigma^2}\right]\right)\right\}\mathrm{d}\tau^2\mathrm{d}\sigma^2, \tag{5.25}$$

and it converges when:

$$\lim_{\sigma^2\to+\infty}\left(\gamma_\sigma^2 - r - r^2\left[\frac{\tau^2}{\sigma^2+n_l\tau^2}+\frac{\sigma^2}{(\sigma^2+n_l\tau^2)^2}\left(\sum_{j=1}^m\frac{n_j}{(\sigma^2+n_j\tau^2)}\right)^{-1}\right]\right) > 0. \tag{5.26}$$

Focusing on the components which involve σ^2, the first term is clearly null, whereas V_μ might be rewritten as:

$$\left(\sum_{j=1}^m\frac{n_j}{\sigma^2+n_j\tau^2}\right)^{-1} = \left(\frac{(\sigma^2)^{m-1}\sum_{j=1}^m\left[n_j\prod_{j'\neq j}\left(1+n_{j'}\tau^2/\sigma^2\right)\right]}{(\sigma^2)^m\prod_{j=1}^m\left(1+n_{j'}\tau^2/\sigma^2\right)}\right)^{-1}. \tag{5.27}$$

Therefore, the limit of the whole addend is:

$$\lim_{\sigma^2\to+\infty}\frac{\sigma^2}{(\sigma^2)^2\left(1+n_l\tau^2/\sigma^2\right)^2}\frac{(\sigma^2)^m\prod_{j=1}^m\left(1+n_j\tau^2/\sigma^2\right)}{(\sigma^2)^{m-1}\sum_{j=1}^m\left[n_j\prod_{j'\neq j}\left(1+n_{j'}\tau^2/\sigma^2\right)\right]} = \frac{1}{n}, \tag{5.28}$$

and the final result is a direct consequence.
(ii) In this case the integral to check is:

$$\begin{aligned}\mathbb{E}\left[\theta_m^r|\mathbf{w}\right] &= \mathbb{E}\left[\exp\left\{r\mu + r\frac{\sigma^2+\tau^2}{2}\right\}|\mathbf{w}\right]\\ &= \int_\Theta \exp\left\{r\mu + r\frac{\sigma^2+\tau^2}{2}\right\}p(\mu|\sigma^2,\tau^2,\mathbf{w})p(\sigma^2,\tau^2|\mathbf{w})\mathrm{d}\boldsymbol{\theta}.\end{aligned} \tag{5.29}$$

Repeating the passages of the previous section, the integral reduces to:

$$\int_0^{+\infty}\int_0^{+\infty} g(\tau^2,\sigma^2)\exp\left\{-\frac{1}{2}\left[\sigma^2\left(\gamma_\sigma^2-r\right)+\tau^2\left(\gamma_\tau^2-r\right)-r^2V_\mu\right]\right\}\mathrm{d}\sigma^2\mathrm{d}\tau^2, \tag{5.30}$$

where $g(\tau^2, \sigma^2)$ does not affect the integral existence. It is possible to collect both the variance components from V_μ, and the finiteness of the integral is assured when the following two conditions hold:

$$\lim_{\sigma^2\to+\infty}\left(\gamma_\sigma^2 - r - r^2\frac{V_\mu}{\sigma^2}\right) > 0, \qquad \lim_{\tau^2\to+\infty}\left(\gamma_\tau^2 - r - r^2\frac{V_\mu}{\tau^2}\right) > 0. \tag{5.31}$$

The first limit is a particular case of the previous result and the same expression is obtained. Focusing on the second one, it is possible to deduce that:

$$\lim_{\tau^2\to+\infty}\frac{(\tau^2)^{m-1}\prod_{j=1}^m\left(\sigma^2/\tau^2+n_j\right)}{(\tau^2)^{m-1}\sum_{j=1}^m\left[n_j\prod_{j'\neq j}\left(\sigma^2/\tau^2+n_{j'}\right)\right]} = \frac{1}{m}; \tag{5.32}$$

and the final result is an immediate consequence. ■

These results have several implications that will be illustrated deeply in the following section about the general model. However, it is evident that if one of the targets of the analysis is the estimation of the local or global mean, then the most common priors for the variance components cannot be used for σ^2 and τ^2. Not only the improper priors like the Jeffrey's one, but also some proper distributions like the inverse gamma or the half-t on the scale parameter (i.e. σ and τ) do not produce posterior distributions with finite moments for θ_m and $\theta_c(\nu_l)$. Therefore, using these priors, the common summary statistics like mean and variance cannot be used to synthesize the samples drawn with usual Monte Carlo techniques from the posteriors.

Moreover, as might be expected, the condition for the existence of the conditional mean functional $\theta_c(\nu_l)$ is equal to the one found by Fabrizi and Trivisano (2012) for the simple log-normal mean. In fact, in this case, the random effects are conditioned and the variance component τ^2 has no restrictions since it does not contribute in increasing the uncertainty of the $\theta_c(\nu_l)$ posterior. On the other hand, considering the global mean, constraints on the τ^2 prior are required to obtain the posterior moments existence. The value of the threshold for γ_τ decreases with the number of groups m in the same way that the threshold on γ_σ decreases with n.

5.1.1 Minimum MSE estimator conditioned to the variance components

In order to have a complete characterization of the estimation problem, a useful finding might be the minimum MSE Bayes estimator, conditioned with respect to the variance components. It is the parallel result of the one by Zellner (1971) for the log-normal mean and of theorem 3.2 for the quantiles. Even if the deduced estimator could be of little practical interest, it might represent an useful benchmark for the considered methods in the simulation study.

For computational easiness, the one-way random effect model (5.2) in the balanced case (i.e. $n_j = n_g,\ \forall j$) is considered. Assuming the variance components σ^2 and τ^2 as known, the only unknown parameter is the global mean in the log-scale μ. Similarly to Zellner (1971), the research of an optimal conditional estimator is restricted to the class of estimators $\theta_m^* = \exp\{\bar{w}\}\, k$. The main result and its relationship with the Bayesian estimation is contained in the following theorem.

Theorem 5.2. *Considering the estimators of the functional θ_m (5.4) that consider σ^2 and τ^2 as known and are included in the class:*

$$\theta_m^* = k \cdot \exp\{\bar{w}\}; \tag{5.33}$$

then the one that minimizes the frequentist MSE is:

$$\hat{\theta}_m^* = \exp\left\{\bar{w} + \frac{\sigma^2+\tau^2}{2} - \frac{3(\sigma^2+n_g\tau^2)}{2n}\right\}. \tag{5.34}$$

Furthermore, it coincides with the conditioned Bayes estimator under the prior $p(\mu) \propto 1$ that minimizes the relative quadratic loss function.

Proof. Recalling that $\bar{w} \sim \mathcal{N}\left(\mu, \sigma^2 n^{-1} + \tau^2 m^{-1}\right)$, the MSE of the considered class of estimator is:

$$\mathbb{E}[(\theta_m^* - \theta_m)^2] = k^2 \exp\left\{2\left(\mu + \sigma^2 n^{-1} + \tau^2 m{-}1\right)\right\} + \\ - 2k \exp\left\{2\mu + \frac{\sigma^2 m + \tau^2 n + nm(\sigma^2 + \tau^2)}{2nm}\right\} + c, \tag{5.35}$$

where c is a constant. The quantity is minimized when:

$$k = \exp\left\{\frac{\sigma^2 + \tau^2}{2} - \frac{3(\sigma^2 + n_g \tau^2)}{2n}\right\}. \tag{5.36}$$

Starting from the (5.17) the expression of the Bayes estimator under relative quadratic loss can be easily derived. ■

For benchmarking purposes, it might be useful to find a minimum MSE estimator conditioned to the variance components for the functional $\theta_c(\nu_j)$ too. In this case, a hard decision to be made is represented by the estimator class restriction, since the global sample mean $\bar{w}$ appears to be not appropriated. A heuristic strategy to obtain a conditioned estimator might be based on the derivation of the Bayes estimator under relative quadratic loss.
Even if it is not proved to be an optimal estimator, its use for benchmarking purposes appears to be largely justified by the good frequentist properties that the Bayes estimator under relative quadratic loss has in the log-normal estimation framework. The formal result is presented in the following proposition.

Proposition 5.1. *The Bayes estimator of $\theta_c(\nu_l)$ conditioned with respect to the variance components under the prior $p(\mu) \propto 1$ that minimizes the relative quadratic loss function is:*

$$\hat{\theta}_c^{RQ}(\nu_l) = \exp\left\{\frac{\sigma^2}{\sigma^2 + n_g \tau^2}\left(\frac{\tau^2 n_g}{\sigma^2}\bar{w}_{.j} - \bar{w}\right) + \frac{\sigma^2}{2} - \frac{3}{2}\frac{\sigma^2}{\sigma^2 + n_g \tau^2}\left(\tau^2 + \frac{\sigma^2}{n}\right)\right\}. \tag{5.37}$$

Proof. To obtain the estimator, the distribution of $\theta_c(\nu_l)|\sigma^2, \tau^2, \mathbf{w}$ must be deduced removing the conditioning on $\boldsymbol{\nu}$ from the (5.18). Setting the value $t_l = \mu + \nu_l$, the following result can be obtained

$$t_l|\sigma^2, \tau^2, \mathbf{w} \sim \mathcal{N}\left(\bar{t}_l, V_{t_l}\right), \tag{5.38}$$

where:

$$\bar{t}_l = \frac{\frac{n_g}{\sigma^2}\bar{w}_{.l}}{\frac{1}{\tau^2} + \frac{n_g}{\sigma^2}} - \frac{\sigma^2}{\sigma^2 + n_g \tau^2}\bar{w}, \qquad V_{t_l} = \frac{\tau^2 \sigma^2}{\sigma^2 + n_g \tau^2} + \left(\frac{\sigma^2}{\sigma^2 + n_g \tau^2}\right)^2 \left(\frac{\sigma^2 + n_g \tau^2}{n}\right). \tag{5.39}$$

Recalling that the Bayes estimator under relative quadratic loss in a log-normal context is:

$$\hat{\theta}_c^{RQ}(\nu_l) = \exp\left\{\bar{t}_l + \frac{\sigma^2}{2} - \frac{3}{2}V_{t_l}\right\}, \tag{5.40}$$

the final result is obtained by substitution. ■

5.2 The log-normal linear mixed model

The one-way random effect model faced in the previous section is a particular case of the more general hierarchical linear mixed model. Considering a vector of responses $\mathbf{y} \in \mathbb{R}^n$, the assumption of log-normality for the response means analysing the log-transformed vector $\mathbf{w} = \log \mathbf{y}$ as normally distributed. The classical formulation of the model is:

$$\mathbf{w} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}. \tag{5.41}$$

The coefficients of the fixed effects are in the vector $\boldsymbol{\beta} \in \mathbb{R}^p$, whereas $\mathbf{u} \in \mathbb{R}^m$ is the vector of random effects and $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the vector of residuals. The design matrices are $\mathbf{X} \in \mathbb{R}^{n\times p}$, that is assumed to be full rank in order to guarantee the existence of $(\mathbf{X}^T\mathbf{X})^{-1}$, and $\mathbf{Z} \in \mathbb{R}^{n\times m}$. The following Bayesian hierarchical model is studied:

$$\begin{aligned} &\mathbf{w}|\mathbf{u}, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{I}_n\sigma^2\right); \\ &\mathbf{u}|\tau_1^2, ..., \tau_q^2 \sim \mathcal{N}_m\left(\mathbf{0}, \mathbf{D}\right), \ \mathbf{D} = \oplus_{s=1}^q \mathbf{I}_{m_s}\tau_s^2; \quad (\boldsymbol{\beta}, \sigma^2) \sim p(\boldsymbol{\beta}, \sigma^2); \\ &\boldsymbol{\tau}^2 \sim p(\tau_1^2, ..., \tau_q^2). \end{aligned} \tag{5.42}$$

Since q random factors are considered, q different variances related to the random components $\boldsymbol{\tau}^2 = (\tau_1^2, ..., \tau_q^2)$ are included in the model. Therefore, it is possible to split the vector of random effects in $\mathbf{u} = [\mathbf{u}_1^T, ..., \mathbf{u}_s^T, ..., \mathbf{u}_q^T]^T$, where $\mathbf{u}_s \in \mathbb{R}^{m_s}$ with $\sum_{s=1}^q m_s = m$. The design matrix of the random effects might be partitioned too: $\mathbf{Z} = [\mathbf{Z}_1 \cdots \mathbf{Z}_s \cdots \mathbf{Z}_q]$. Moreover, it is worth to highlight that the design matrix of the random effects is not necessarily non-singular. Hence, to invert the quantity $\mathbf{Z}^T\mathbf{Z}$, the Moore-Penrose inverse might be required. Finally, the covariance matrix of the random effects $\mathbf{D}$ is assumed to be diagonal, but with a simple linear transformation it is possible to consider a block diagonal matrix $\oplus_{s=1}^q \mathbf{G}_s\tau_s^2$, where $\mathbf{G}_s \in \mathbb{R}^{m_s\times m_s}$ is a fixed positive definite matrix (Hobert and Casella, 1996).
This model formulation is particularly general and it is widely used with a log-transformed response in many applied fields: actuarial sciences (Antonio et al., 2006); occupational health (Lyles et al., 1997a,b; Krishnamoorthy and Mathew, 2002); medicine (Coursaget et al., 1991; Berchialla et al., 2009), psychology (Van Breukelen, 2005), environmental sciences and ecology (Price et al., 1995; Hector et al., 2012), without ignoring the small area estimation framework (Berg and Chandra, 2014; Molina and Martin, 2018).
In order to obtain meaningful and explainable results, it is often necessary to compute the estimates of the expectation conditioned with respect to either the fixed and random effects or conditioned only to a fixed covariate pattern (i.e. integrating out the random effects). Therefore, in practice, the interpretable outputs are usually provided in the original data scale, back-transforming the results obtained with the model (5.42). It is well known that, exploiting the properties of the log-normal distribution, the conditioned expectation of the observation $\tilde{y}$ given the random effects and the covariate patterns $\tilde{\mathbf{x}}$, $\tilde{\mathbf{z}}$ (quantity that could be also labelled as subject-specific expectation) is:

$$\mathbb{E}\left[\tilde{y}|\mathbf{u}, \boldsymbol{\beta}, \sigma^2\right] = \theta_c(\tilde{\mathbf{x}}, \tilde{\mathbf{z}}) = \exp\left\{\tilde{\mathbf{x}}^T\boldsymbol{\beta} + \tilde{\mathbf{z}}^T\mathbf{u} + \frac{\sigma^2}{2}\right\}; \tag{5.43}$$

whereas, if the random effects are ignored and they are integrated out, then the conditioned expectation of interest is:

$$\mathbb{E}\left[\tilde{y}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\tau}\right] = \theta_m(\tilde{\mathbf{x}}) = \exp\left\{\tilde{\mathbf{x}}^T\boldsymbol{\beta} + \frac{1}{2}\left(\sigma^2 + \sum_{s=1}^{q}\tau_s^2\right)\right\}. \tag{5.44}$$

The posterior predictive distribution $p(\tilde{y}|\mathbf{y})$ and its posterior moments is a further quantity that might be investigated. It is obtained exponentiating the posterior predictive distribution of the variable $\tilde{w} = \log\tilde{y}$, having as covariate patterns $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{z}}$. In general, it is defined as:

$$p(\tilde{y}|\mathbf{y},\tilde{\mathbf{x}},\tilde{\mathbf{z}}) \propto \int_{\boldsymbol{\Theta}} p(\tilde{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})\mathrm{d}\boldsymbol{\theta}, \tag{5.45}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta},\mathbf{u},\sigma^2,\boldsymbol{\tau}^2)$ and Θ is the parameter space. Moreover, the distribution of the transformed response variable conditioned on the parameters is:

$$\tilde{w}|\boldsymbol{\beta},\mathbf{u},\sigma^2,\boldsymbol{\tau}^2 \sim \mathcal{N}\left(\tilde{\mathbf{x}}^T\boldsymbol{\beta} + \tilde{\mathbf{z}}^T\mathbf{u},\sigma^2\right). \tag{5.46}$$

In practice, the posterior expectation $\mathbb{E}\left[\tilde{y}|\mathbf{y}\right]$ might be used to predict unobserved values like missing values or unsampled units in small area estimation.
In many applications placed in the Bayesian context, the estimation of the target quantity is performed through the mean of their posterior distributions. However, it will be shown that a careful specification of the priors for the variance components is necessary, in order to assure the existence of the posterior moments of the functionals $\theta_c(\tilde{\mathbf{x}},\tilde{\mathbf{z}})$, $\theta_m(\tilde{\mathbf{x}})$ and of the predictive distribution.
Likewise the one-way random effect ANOVA model explained in the previous section, a flat improper prior $p(\boldsymbol{\beta}) \propto 1$ is considered for the vector of coefficients $\boldsymbol{\beta}$ related to the fixed effects. On the other hand, independent GIG priors are adopted for the variance components:

$$p(\sigma^2) \sim GIG(\lambda_\sigma,\delta_\sigma,\gamma_\sigma); \quad p(\tau_s^2) \sim GIG(\lambda_{\tau,s},\delta_{\tau,s},\gamma_{\tau,s}),\ \forall s. \tag{5.47}$$

As a consequence, in this setting, the joint posterior distribution for the model is:

$$\begin{aligned} p(\boldsymbol{\beta},\mathbf{u},\sigma^2,\boldsymbol{\tau}^2|\mathbf{w}) \propto p(\sigma^2)\left(\prod_{s=1}^{q}p(\tau_s^2)\right)|\mathbf{I}_n\sigma^2|^{\frac{1}{2}}|\mathbf{D}|^{-\frac{1}{2}}\times \\ \times\exp\left\{-\frac{1}{2\sigma^2}\left(\mathbf{w}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}\right)^T\left(\mathbf{w}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}\right)\right\}\times \\ \times\exp\left\{-\frac{1}{2}\mathbf{u}^T\mathbf{D}^{-1}\mathbf{u}\right\}. \end{aligned} \tag{5.48}$$

It is possible to exploit the conditioned-conjugacy of the GIG priors in order to derive the full conditionals:

$$\sigma^2|\boldsymbol{\beta},\mathbf{u},\boldsymbol{\tau}^2,\mathbf{w} \sim GIG\left(\lambda_\sigma - \frac{n}{2}, \sqrt{(\mathbf{w}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^T(\mathbf{w}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})+\delta_\sigma^2}, \gamma_\sigma\right); \quad (5.49)$$

$$\tau_s^2|\boldsymbol{\beta},\mathbf{u},\sigma^2,\boldsymbol{\tau}_{-s}^2,\mathbf{w} \sim GIG\left(\lambda_{\tau,s} - \frac{m_s}{2}, \sqrt{\mathbf{u}_s^T\mathbf{u}_s+\delta_{\tau,s}^2}, \gamma_{\tau,s}\right), \qquad s=1,...,q; \quad (5.50)$$

$$\mathbf{u}|\boldsymbol{\beta},\sigma^2,\boldsymbol{\tau}^2,\mathbf{w} \sim \mathcal{N}_m\left(\mathbf{V_u}\mathbf{Z}^T(\mathbf{w}-\mathbf{X}\boldsymbol{\beta}),\sigma^2\mathbf{V_u}\right); \quad (5.51)$$

$$\boldsymbol{\beta}|\mathbf{u},\sigma^2,\boldsymbol{\tau}^2,\mathbf{w} \sim \mathcal{N}_p\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T(\mathbf{w}-\mathbf{Z}\mathbf{u}),\sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right); \quad (5.52)$$

where $\mathbf{V_u} = \left(\mathbf{Z}^T\mathbf{Z}+\sigma^2\mathbf{D}^{-1}\right)^{-1}$.
In order to better characterize the posterior properties of the model and to prove the main results of this section about the posterior moments of interest, some useful distributions are derived in the following proposition.

Proposition 5.2. *Considering the log-normal linear mixed model* (5.42) *with a flat improper prior on* $\boldsymbol{\beta}$ *and the independent GIG priors* (5.47) *for the variance components, then, starting from the joint posterior* (5.48), *it is possible to deduce the following distributional results:*

(i)

$$\boldsymbol{\beta}|\sigma^2,\boldsymbol{\tau}^2,\mathbf{w} \sim \mathcal{N}_p\left(\bar{\boldsymbol{\beta}},\mathbf{V}_\beta\right), \quad (5.53)$$

where:

$$\mathbf{V}_\beta = \left(\frac{(\mathbf{X}^T\mathbf{X})}{\sigma^2}+\mathbf{X}^T\mathbf{M}\mathbf{X}\right)^{-1}, \quad \bar{\boldsymbol{\beta}} = \mathbf{V}_\beta\left(\frac{\mathbf{X}^T\mathbf{X}}{\sigma^2}\hat{\boldsymbol{\beta}}+\mathbf{X}^T\mathbf{M}\mathbf{X}\tilde{\boldsymbol{\beta}}\right),$$
$$\mathbf{M} = \left(\frac{\mathbf{V}_Z^{-1}}{\sigma^2}-\frac{\mathbf{P}_Z}{\sigma^2}\right), \quad \hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}, \quad \tilde{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{M}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{M}\mathbf{y}, \quad (5.54)$$
$$\mathbf{P}_Z = \mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^{-}\mathbf{Z}^T, \quad \mathbf{V}_Z^{-1} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-}\left((\mathbf{Z}^T\mathbf{Z})^{-}+\frac{\mathbf{D}}{\sigma^2}\right)^{-1}(\mathbf{Z}^T\mathbf{Z})^{-}\mathbf{Z}^T,$$

and $\left(\mathbf{Z}^T\mathbf{Z}\right)^{-}$ *is the Moore-Penrose inverse of* $\mathbf{Z}^T\mathbf{Z}$.

(ii)

$$\begin{aligned} p(\sigma^2,\boldsymbol{\tau}^2|\mathbf{w}) \propto\ & p(\sigma^2)\left(\prod_{s=1}^r p(\tau_s^2)\right)(\sigma^2)^{-\frac{n-m}{2}}|\sigma^2(\mathbf{Z}^T\mathbf{Z})^{-}+\mathbf{D}|^{\frac{1}{2}}\times \\ & \times\left|\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}+\left(\mathbf{X}^T\mathbf{M}\mathbf{X}\right)^{-1}\right|^{-\frac{1}{2}}\times \\ & \times\exp\left\{-\frac{1}{2}\left[\frac{RSS}{\sigma^2}+\left(\mathbf{w}-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)^T\mathbf{M}\left(\mathbf{w}-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)+\right.\right. \\ & \left.\left.\times\left(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)^T\left(\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}+\left(\mathbf{X}^T\mathbf{M}\mathbf{X}\right)^{-1}\right)\left(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)\right]\right\}, \end{aligned} \quad (5.55)$$

where:

$$RSS = \left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^T\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right). \quad (5.56)$$

Proof. The matrix algebra passages used to prove the statements of this proposition are based on the following well known equalities involving quadratic forms:

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{u}) = {} & (\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+\\ & -(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{Z}\left(\mathbf{Z}^T\mathbf{Z}\right)^-\mathbf{Z}^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})+\\ & +\left(\mathbf{u}-\left(\mathbf{Z}^T\mathbf{Z}\right)^-\mathbf{Z}^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right)^T\mathbf{Z}^T\mathbf{Z}\times\\ & \times\left(\mathbf{u}-\left(\mathbf{Z}^T\mathbf{Z}\right)^-\mathbf{Z}^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right),\end{aligned} \tag{5.57}$$

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)^T\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)+(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})^T\mathbf{X}^T\mathbf{X}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}), \tag{5.58}$$

$$\begin{aligned}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^T\mathbf{S}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = {} & \left(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)^T\mathbf{S}^{-1}\left(\mathbf{y}-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)+\\ & +(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^T\mathbf{X}^T\mathbf{S}^{-1}\mathbf{X}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}).\end{aligned} \tag{5.59}$$

Moreover, the following result about the product of two d-dimensional normal densities is employed:

$$\begin{aligned}\mathcal{N}_d(\mathbf{x}|\mathbf{a},\mathbf{A})\,\mathcal{N}_d(\mathbf{x}|\mathbf{b},\mathbf{B}) = {} & \mathcal{N}_d(\mathbf{x}|\mathbf{c},\mathbf{C})\,(2\pi)^{-\frac{d}{2}}|\mathbf{A}+\mathbf{B}|^{-\frac{1}{2}}\times\\ & \times\exp\left\{-\frac{1}{2}(\mathbf{a}-\mathbf{b})^T(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{a}-\mathbf{b})\right\},\end{aligned} \tag{5.60}$$

where $\mathbf{c} = \mathbf{C}\left(\mathbf{A}^{-1}\mathbf{a}+\mathbf{B}^{-1}\mathbf{b}\right)$ and $\mathbf{C} = \left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}$.
In order to deduce the result (i), the likelihood marginalized with respect to the random effects can be easily obtained:

$$\begin{aligned}L(\boldsymbol{\beta},\boldsymbol{\tau}^2,\sigma^2) & = \int_{\mathbb{R}^m} p(\mathbf{w}|\mathbf{u},\boldsymbol{\beta},\sigma^2)p(\mathbf{u}|\boldsymbol{\tau}^2)\mathrm{d}\mathbf{u}\\ & \propto (\sigma^2)^{-\frac{n-m}{2}}|\sigma^2(\mathbf{Z}^T\mathbf{Z})^-+\mathbf{D}|^{\frac{1}{2}}\times\\ & \times\exp\left\{-\frac{1}{2\sigma^2}\left[RSS+(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})^T(\mathbf{X}^T\mathbf{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})\right]\right\}\times\\ & \times\exp\left\{-\frac{1}{2}\left[\left(\mathbf{w}-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)^T\mathbf{M}\left(\mathbf{w}-\mathbf{X}\tilde{\boldsymbol{\beta}}\right)+\right.\right.\\ & \left.\left.\quad+\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T\mathbf{X}^T\mathbf{M}\mathbf{X}\left(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\right]\right\}.\end{aligned} \tag{5.61}$$

Then, the (5.53) follows through a simple application of the (5.60).
(ii) As far as the result (5.55) it concerns, it might be derived from the marginalized likelihood (5.61) adding the priors (5.47) and integrating out $\boldsymbol{\beta}$. ■

Before stating the main result of the chapter, an useful quantity to define is the matrix $\mathbf{L}_s \in \mathbb{R}^{p\times p}$: its entries are all 0s with the exception of the first $l \times l$ square block $\mathbf{L}_{s;1,1}$, where $l = p - \mathrm{rank}\{\mathbf{X}^T(\mathbf{I}-\mathbf{P}_\mathbf{Z})\mathbf{X}\}$. It coincides with the number of variables of $\mathbf{X}$ that are

included in $\mathbf{Z}$ too. Furthermore, to simplify the final form of the result, it is useful to place the columns related to these variables as first l columns of the *ordered design matrix* $\mathbf{X}_o$, without loss of generality. As a consequence, the matrix $\mathbf{L}_{s;1,1}$ coincides with the inverse of the upper left $l \times l$ block on the diagonal of the matrix $\mathbf{X}_o^T \left(\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-}\mathbf{C}_s(\mathbf{Z}^T\mathbf{Z})^{-}\mathbf{Z}^T\right)\mathbf{X}_o$, where $\mathbf{C}_s$ is the null matrix with the exception of $\mathbf{I}_{m_s}$ as block on the diagonal in correspondence to the s-th variance component of the random effect. To complete the notation, $\tilde{\mathbf{x}}_o$ is the covariate pattern of the observation to estimate that is ordered coherently with respect to $\mathbf{X}_o$.
The results of the previous proposition can be used in order to prove the main findings contained in the following theorem. It states the conditions to fulfil in the prior specification step in order to have the finiteness of the posterior moments for the functionals (5.43), (5.44) and for the posterior predictive distribution.

Theorem 5.3. *If the normal linear mixed model in the log-scale* (5.42) *is considered with the priors* (5.47)*, then, in order to compute the r-th posterior moment (with $r > 0$) of $\theta_c(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$, $\theta_m(\tilde{\mathbf{x}})$ and of $p(\tilde{y}|\mathbf{y})$, the following constraints on the prior parameters must be fulfilled:*

(i) $\mathbb{E}\left[\theta_c^r(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})|\mathbf{w}\right]$ *exists if* $\gamma_\sigma^2 > r + r^2\tilde{\mathbf{x}}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\tilde{\mathbf{x}}$*;*

(ii) $\mathbb{E}\left[\theta_m^r(\tilde{\mathbf{x}})|\mathbf{w}\right]$ *exists if* $\gamma_\sigma^2 > r + r^2\tilde{\mathbf{x}}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\tilde{\mathbf{x}}$ *and* $\gamma_{\tau,s}^2 > r + r^2\tilde{\mathbf{x}}_o^T\mathbf{L}_s\tilde{\mathbf{x}}_o,\ \forall s$*;*

(iii) $\mathbb{E}\left[\tilde{y}^r|\mathbf{y}\right]$ *exists if* $\gamma_\sigma^2 > r^2 + r^2\tilde{\mathbf{x}}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\tilde{\mathbf{x}}$.

Proof. (i) Proceeding similarly with respect to the proof of theorem 5.1, the r-th moment of $\theta_c(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$ is defined as:

$$\begin{aligned}\mathbb{E}\left[\theta_c^r(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})|\mathbf{w}\right] &= \mathbb{E}\left[\exp\left\{r\tilde{\mathbf{x}}^T\boldsymbol{\beta} + r\tilde{\mathbf{z}}^T\mathbf{u} + r\frac{\sigma^2}{2}\right\}|\mathbf{w}\right] \\ &= \int_{\boldsymbol{\Theta}} \exp\left\{r\tilde{\mathbf{x}}^T\boldsymbol{\beta} + r\tilde{\mathbf{z}}^T\mathbf{u} + r\frac{\sigma^2}{2}\right\} p(\boldsymbol{\beta}, \mathbf{u}, \sigma^2, \boldsymbol{\tau}^2|\mathbf{w})\mathrm{d}\boldsymbol{\theta}.\end{aligned} \tag{5.62}$$

Recalling the (5.51), (5.53) and the (5.55) and performing a simple change of variable, it is possible to solve the integral twice recognizing the moment generating function of a Gaussian distribution: the first related to $\mathcal{N}\left(\tilde{\mathbf{z}}^T\mathbf{V}_\mathbf{u}\mathbf{Z}^T\left(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}\right), \sigma^2\tilde{\mathbf{z}}^T\mathbf{V}_\mathbf{u}\tilde{\mathbf{z}}\right)$ and the second to $\mathcal{N}_p\left(\tilde{\mathbf{q}}^T\bar{\boldsymbol{\beta}}, \tilde{\mathbf{q}}^T\mathbf{V}_\beta\tilde{\mathbf{q}}\right)$, where $\tilde{\mathbf{q}}^T = \tilde{\mathbf{z}}^T\mathbf{V}_\mathbf{u}\mathbf{Z}^T\mathbf{X} - \tilde{\mathbf{x}}^T$. Then, the following integral is obtained:

$$\begin{aligned}\int_0^{+\infty}\cdots\int_0^{+\infty} g(\sigma^2, \boldsymbol{\tau}^2)\exp\Bigg\{-\frac{1}{2}\sigma^2\Big(\gamma_\sigma^2 - r + \\ -r^2\left[\tilde{\mathbf{z}}^T\mathbf{V}_\mathbf{u}\tilde{\mathbf{z}} + \frac{\tilde{\mathbf{q}}^T\mathbf{V}_\beta\tilde{\mathbf{q}}}{\sigma^2}\right]\Big)\Bigg\}\mathrm{d}\boldsymbol{\tau}^2\mathrm{d}\sigma^2,\end{aligned} \tag{5.63}$$

where $g(\sigma^2, \boldsymbol{\tau}^2)$ is a functional of the variance components that does not affect the finiteness of the integral. Therefore, the integral is finite when:

$$\lim_{\sigma^2\to+\infty}\left(\gamma_\sigma^2 - r - r^2\left[\tilde{\mathbf{z}}^T\mathbf{V}_\mathbf{u}\tilde{\mathbf{z}} + \frac{\tilde{\mathbf{q}}^T\mathbf{V}_\beta\tilde{\mathbf{q}}}{\sigma^2}\right]\right) > 0. \tag{5.64}$$

In order to compute this limit, lemma 1 by Hobert and Casella (1996) is useful. It states that, given a scalar c and a non-negative definite matrix $\mathbf{S}$, the limit:

$$\lim_{c \to +\infty} \left(\mathbf{S} + \frac{\mathbf{I}}{c} \right)^{-1}, \tag{5.65}$$

coincides with a generalized inverse of $\mathbf{S}$. Moreover, it is immediate to extend the result to the case in which any diagonal matrix substitutes $\mathbf{I}$.
Considering the limit of the factor that multiplies r^2 in the (5.64) and focusing on the first addend, recalling that $\mathbf{V_u} = \left(\mathbf{Z}^T\mathbf{Z} + \sigma^2\mathbf{D}^{-1}\right)^{-1}$, by applying the previous result and doing some computations, it is possible to show that:

$$\lim_{\sigma^2 \to +\infty} \tilde{\mathbf{z}}^T \left(\mathbf{Z}^T\mathbf{Z} + \sigma^2\mathbf{D}^{-1}\right)^{-1} \tilde{\mathbf{z}} = \lim_{\sigma^2 \to +\infty} \frac{1}{\sigma^2} \tilde{\mathbf{z}}^T \left(\frac{\mathbf{Z}^T\mathbf{Z}}{\sigma^2} + \mathbf{D}^{-1} \right)^{-1} \tilde{\mathbf{z}} = 0. \tag{5.66}$$

Then, the limit of the second added must be computed. It is:

$$\lim_{\sigma^2 \to +\infty} \frac{\tilde{\mathbf{q}}^T \mathbf{V}_\beta \tilde{\mathbf{q}}}{\sigma^2} = \lim_{\sigma^2 \to +\infty} \tilde{\mathbf{q}}^T \left(\mathbf{X}^T\mathbf{X} + \sigma^2\mathbf{X}^T\mathbf{M}\mathbf{X}\right)^{-1} \tilde{\mathbf{q}}. \tag{5.67}$$

Focusing on the structure of the matrix $\mathbf{M}$:

$$\sigma^2\mathbf{X}^T\mathbf{M}\mathbf{X} = \mathbf{X}^T \left(\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^- \left((\mathbf{Z}^T\mathbf{Z})^- + \frac{\mathbf{D}}{\sigma^2} \right)^{-1} (\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T \right) \mathbf{X}, \tag{5.68}$$

and using the result on limit (5.65), it can be noted that:

$$\lim_{\sigma^2 \to +\infty} \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^- \left((\mathbf{Z}^T\mathbf{Z})^- + \frac{\mathbf{D}}{\sigma^2} \right)^{-1} (\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T. \tag{5.69}$$

Therefore, it is possible to conclude that the limit reduces to:

$$\lim_{\sigma^2 \to +\infty} \tilde{\mathbf{q}}^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{q}} = \lim_{\sigma^2 \to +\infty} \left(\tilde{\mathbf{z}}^T\mathbf{V_u}\mathbf{Z}^T\mathbf{X} - \tilde{\mathbf{x}}^T\right) \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \left(\tilde{\mathbf{z}}^T\mathbf{V_u}\mathbf{Z}^T\mathbf{X} - \tilde{\mathbf{x}}^T\right)^T. \tag{5.70}$$

Hence, solving the deduced quadratic form and computing the limits similarly to (5.66), it is finally obtained the result:

$$\lim_{\sigma^2 \to +\infty} \frac{\tilde{\mathbf{q}}^T \mathbf{V}_\beta \tilde{\mathbf{q}}}{\sigma^2} = \tilde{\mathbf{x}}^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}. \tag{5.71}$$

The concluding algebraic passages are straightforward.
(ii) In this case, the integral defining the r-th posterior moment of $\theta_m(\tilde{\mathbf{x}})$ might be decomposed as:

$$\begin{aligned} \mathbb{E}\left[\theta_c^r(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})|\mathbf{w}\right] &= \mathbb{E}\left[\exp\left\{ r\tilde{\mathbf{x}}^T\boldsymbol{\beta} + \frac{r}{2}\left(\sigma^2 + \sum_{s=1}^q \tau_s^2\right)\right\} |\mathbf{w} \right] \\ &= \int_0^{+\infty} \cdots \int_0^{+\infty} g(\sigma^2, \boldsymbol{\tau}^2) \exp\left\{ -\frac{1}{2}\left[\sigma^2(\gamma_\sigma^2 - r) + \right.\right. \\ &\qquad \left.\left. + \sum_{s=1}^r \tau_s^2(\gamma_{\tau,s}^2 - r) - r^2\tilde{\mathbf{x}}^T\mathbf{V}_\beta\tilde{\mathbf{x}} \right]\right\} \mathrm{d}\sigma^2 \mathrm{d}\boldsymbol{\tau}^2. \end{aligned} \tag{5.72}$$

In order to verify the finiteness of the previous integral, the behaviour of term $r^2\tilde{\mathbf{x}}^T\mathbf{V}_\beta\tilde{\mathbf{x}}$ must be checked when all the variance components go to $+\infty$.
The limit for $\sigma^2 \to +\infty$ gives the same result of point (i), whereas the limit for the generic term τ_s^2 can be written as:

$$\lim_{\tau_s^2\to+\infty} \sigma^2\tilde{\mathbf{x}}^T\Bigg[\tau_s^2\mathbf{X}^T\left(\mathbf{I}-\mathbf{P}_\mathbf{Z}\right)\mathbf{X}+ +\mathbf{X}^T\left(\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^-\left(\frac{(\mathbf{Z}^T\mathbf{Z})^-}{\tau_s^2}+\frac{\mathbf{D}}{\tau_s^2\sigma^2}\right)^{-1}(\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T\right)\mathbf{X}\Bigg]^{-1}\tilde{\mathbf{x}}. \tag{5.73}$$

By taking the limit $\tau_s^2 \to +\infty$ to the term $\left(\frac{(\mathbf{Z}^T\mathbf{Z})^-}{\tau_s^2}+\frac{\mathbf{D}}{\tau_s^2\sigma_s^2}\right)$, a matrix $\mathbf{C}_s$ is obtained. All its elements are null with the exception of the presence of $\mathbf{I}_{m_s}$ as block on the diagonal in correspondence to the s-th variance component related to the vector of random effects $\mathbf{u}_s$ in matrix $\mathbf{D}$. Moreover, its generalized inverse is the matrix $\mathbf{C}_s$ itself. Therefore, the limit can be written as:

$$\lim_{\tau_s^2\to+\infty} \sigma^2\tilde{\mathbf{x}}_o^T\left[\tau_s^2\mathbf{X}_o^T\left(\mathbf{I}-\mathbf{P}_\mathbf{Z}\right)\mathbf{X}_o+\sigma^2\mathbf{X}_o^T\left(\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^-\mathbf{C}_s(\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T\right)\mathbf{X}_o\right]^{-1}\tilde{\mathbf{x}}_o, \tag{5.74}$$

where $\mathbf{X}$ and $\tilde{\mathbf{x}}$ have been respectively replaced by their ordered versions $\mathbf{X}_o$ and $\tilde{\mathbf{x}}_o$, without loss of generality. Thanks to this ordered matrix, the first term $\mathbf{A} = \mathbf{X}_o^T\left(\mathbf{I}-\mathbf{P}_\mathbf{Z}\right)\mathbf{X}_o$ can be written as:

$$\tau_s^2\mathbf{A} = \begin{bmatrix} \mathbf{0}_l & \mathbf{0}^T \\ \mathbf{0} & \tau_s^2\mathbf{A}_{2,2} \end{bmatrix}, \tag{5.75}$$

where $\mathbf{0}_l$ is the null squared matrix of dimension l, that is the rank deficiency of $\mathbf{A}$. This feature is due to the ordering of $\mathbf{X}_o$ and the linear algebraic dependence of the first l columns of $\mathbf{X}_o$ to the columns of $\mathbf{Z}$. Denoting with $\mathbf{B}_s$ the second matrix, then the sum $\mathbf{A}+\mathbf{B}$ can be written as:

$$\begin{bmatrix} \mathbf{B}_{s;1,1} & \boldsymbol{B}_{s;1,2}^T \\ \boldsymbol{B}_{s;1,2} & \tau^2\mathbf{A}_{2,2}+\mathbf{B}_{s;2,2} \end{bmatrix}. \tag{5.76}$$

To complete the proof, the result of the limit can be written as:

$$\tilde{\mathbf{x}}_o^T\mathbf{L}_s\tilde{\mathbf{x}}_o. \tag{5.77}$$

In addition, it can be noted that exploiting the property of the block matrix inverse:

$$\mathbf{L}_s = \begin{bmatrix} \mathbf{L}_{s;1,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{p-l} \end{bmatrix}, \tag{5.78}$$

and $\mathbf{L}_{s;1,1} = \mathbf{B}_{s;1,1}^{-1} \in \mathbb{R}^{l\times l}$.
(iii) Recalling the definitions of the posterior predictive distribution (5.45) and (5.46), the moments of interest might be defined as:

$$\mathbb{E}\left[\tilde{y}^r|\mathbf{y}\right] = \int_\Theta\left(\int_{-\infty}^{+\infty}\exp\{r\tilde{w}\}\,p(\tilde{w}|\boldsymbol{\beta},\mathbf{u},\sigma^2)\mathrm{d}\tilde{w}\right)p(\mathbf{u},\boldsymbol{\beta},\sigma^2,\boldsymbol{\tau}^2|\mathbf{y})\mathrm{d}\boldsymbol{\theta}. \tag{5.79}$$

Following algebraic passages similar to the proof of (i) the final result is obtained. ■

The results included in the previous general theorem could raise various considerations and some of them were already hinted in section 5.1. As a first instance, coherently with the results by Fabrizi and Trivisano (2012), Fabrizi and Trivisano (2016) and the first part of this work on the log-normal quantiles, the tail parameter γ of the GIG distribution is the one to be constrained in order to determine the existence of posterior moments. Moreover, the only functional that requires a condition on the priors of the whole set of variances is $\theta_m(\tilde{\mathbf{x}})$. In fact, both the subject specific predictor $\theta_c(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$ and the posterior predictive distribution necessitate only a single condition on the variance term σ^2 to ensure the finiteness of the posterior moments.
Focusing on condition (i), it perfectly matches the result by Fabrizi and Trivisano (2016) for the log-normal linear model: the square of the moment order r is multiplied by the leverage $\tilde{\mathbf{x}}^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}$. Concerning this result, it is worth to point out that the condition found for the one-way ANOVA model in theorem 5.1 and, more generally, for the case $\mathbf{X} = \mathbf{1}$, is a direct consequence of this result, since the leverage is equal to n^{-1}. The same condition on γ_σ appears also when the interest is on the moments of $\theta_m(\tilde{\mathbf{x}})$. However, in that case, the priors on τ_s^2 require a constraint on the tail parameters too. In particular, the form of the second result in statement (ii) appear to be rather complicated, mainly due to its generality. To better understand the condition, a couple of common formulations for the matrix $\mathbf{Z}$ might be analysed stressing that, if a unique random effect is considered, then $\mathbf{C}_s = \mathbf{I}_m$. For example, when $\mathbf{Z}$ is built in order to express a random intercept, the problem largely simplifies. In this case, the numbers of columns m coincides with the number of groups of data and for each subject a 1 is marked in the corresponding column. Consequently, $\mathbf{X}_o$ is the simple design matrix since the first column is the usual vector $\mathbf{1}$ describing the intercept and the first element of $\tilde{\mathbf{x}}_o$ is clearly 1. Moreover, it is easy to verify that the rank deficiency of $\mathbf{X}^T (\mathbf{I} - \mathbf{P}_\mathbf{Z}) \mathbf{X}$ is $l = 1$ and therefore the unique non-null entry of $\mathbf{L}_s$ is the first element of the first column. Finally, exploiting the particular structure of $\mathbf{Z}$, after some algebra it is possible to verify that $\mathbf{L}_{s;1,1} = m^{-1}$ (e.g. the inverse of the number of groups determined by $\mathbf{Z}$). In this way the result (ii) of theorem 5.1 on the one-way ANOVA model is obtained too, as a special case.
Another interesting instance to investigate is the case in which $\mathbf{Z}$ identifies a random coefficient: the matrix is built preserving the same structure of the random intercept but, instead of 1 the matrix is filled with the values of the covariate $\mathbf{X}_{,k}$ of interest. Note that $\mathbf{X}_{,k}$ represents the generic k-th column of the matrix $\mathbf{X}$ and is the first column of the ordered matrix $\mathbf{X}_o$, since it is linearly dependent to the columns of $\mathbf{Z}$. In this condition, $\tilde{\mathbf{x}}_o^T \mathbf{L}_s \tilde{\mathbf{x}}_o$ could be interpreted as a sort of leverage weighted with respect to the group structure.
On the other hand, if $q > 1$ distinct random effects are included in the model, then the generic formulation with $\mathbf{C}_s$ is required and the meaning of the result becomes less intuitive.
Finally, as far as the posterior predictive distribution it concerns, the existence of its posterior moments is related only to the term σ^2. It must be noted that, unlike case (i), the term r^2 directly enters the condition, making rapidly increase the value of the constraint with the moment order. The result is in accordance with the higher variability that characterizes the posterior predictive distribution, if compared to the posterior of the functionals $\theta_c(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$ and

$\theta_m(\tilde{\mathbf{x}})$.
Furthermore, a crucial consequence of theorem 5.3 is that many priors for variance components which are used in the literature are not appropriate if the statistical analysis include the computation of the posterior moments considered in this work. The following proposition synthesizes which prior $p(\omega^2)$ could be adopted and which ones should be avoided for the generic variance component ω^2.

Proposition 5.3. *Considering the log-normal linear mixed model* (5.42) *and the results of theorem 5.3, if a condition on the variance component prior is necessary, then the following common priors cannot be used: improper priors (e.g. Jeffrey's or uniform), inverse gamma, log-normal, half-t (included half-Cauchy), and, more generally, all the priors on the scale parameter ω. On the other hand, the following distributions might be chosen, observing the hyperparameter condition: the GIG distribution and the half-normal distribution (on ω^2).*

Proof. From the proof of theorem 5.3, it appears that the prior on the generic variance component $p(\omega^2)$, in order to preserve the existence of the moments, must contain the following exponential term: $\exp\left\{-c\omega^2\right\}$, where c is an hyperparameter (or a function of it) that can be fixed according to the derived conditions. ■

Intuitively, all the distributions that do not have an exponential term in the density function do not preserves the existence of the moments of interest. Moreover, distributions like the inverse gamma or the log-normal, even if an exponential is present, cannot be used since they do not compensate the explosiveness of the term in ω^2 when $\omega^2 \to +\infty$. Once a suitable distribution is chosen, then the constraints on the hyperparameters must be checked. This work considers the GIG distribution as prior and therefore the meaning of the constraints on its hyperparameters have been commented already. Anyway, it must be highlighted that it has as special cases also the exponential, gamma and inverse Gaussian distributions that might represent proper choices. On the other hand, the inverse gamma is a limiting distribution too, but when the tail parameter γ goes to 0, so it is excluded by the existence condition.
As showed in proposition 5.3, another distribution that could be considered as prior for the variance components is the half-normal. It is mentioned as reasonable prior for the scale with a large scale parameter V_ω in the well-known paper by Gelman (2006):

$$\omega|V_\omega \sim \text{half-}\mathcal{N}\left(0, V_\omega\right). \tag{5.80}$$

However, to get finite posterior moments, the half normal must be specified as prior for the variance itself, taking into account the constraints of theorem 5.3. Hence, the scale parameter of the prior should be fixed lower than the square root of the reciprocal of the conditions for γ^2. As an example the condition of point (i) would be:

$$V_\omega < \sqrt{\frac{1}{2\left[r + r^2\tilde{\mathbf{x}}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\tilde{\mathbf{x}}\right]}}. \tag{5.81}$$

Intuitively, the tail decay of such a prior might be too rapid and an excessive amount of prior information might be included in the model, whereas the GIG distribution provides useful tools to control it and to specify a more suitable prior distribution.
Until now, the case of a single point estimation with covariate patterns $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{z}}$ is considered. More generally, $\tilde{n}$ covariate patterns can be stored in the matrices $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Z}}$, therefore the existence conditions of theorem 5.3 can be immediately generalized as follows.

Proposition 5.4. *Considering the model considered in theorem 5.3 and $\tilde{n}$ points to estimate, then the conditions (i), (ii) and (iii) can be generalized as:*

(i) $\mathbb{E}\left[\theta_c^r(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})|\mathbf{w}\right]$ *exists if* $\gamma_\sigma^2 > r + r^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}_i\right)$;

(ii) $\mathbb{E}\left[\theta_m^r(\tilde{\mathbf{x}})|\mathbf{w}\right]$ *exists if* $\gamma_\sigma^2 > r + r^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}_i\right)$ *and, for the variances of the random effects:* $\gamma_{\tau,s}^2 > r + r^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_{o,i}^T \mathbf{L}_s \tilde{\mathbf{x}}_{o_i}\right), \ \forall s$;

(iii) $\mathbb{E}\left[\tilde{y}^r|\mathbf{y}\right]$ *exists if* $\gamma_\sigma^2 > r^2 + r^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}_i\right)$.

5.2.1 The Gibbs sampler and software details

As pointed out by Chib and Carlin (1999), the mere use of the full conditionals to build the Gibbs sampler might lead to autocorrelation issues in the samples from the posterior of $\boldsymbol{\beta}$. To overcome this issue, they proposed the use of their algorithm 2, which is based on the generation of $\boldsymbol{\beta}$ from the posterior conditioned with respect to the variance components (5.53) only, whereas the vector of random effects $\mathbf{u}$ is integrated out. In this way, it is possible to update the vectors of fixed and random effects a the same time, preserving the structure of Gibbs sampler:

1. Initialize the vector of parameters $(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}^2, \sigma^2)$;
2. Repeat for the desired number of iteration:
   (a) Sample $\boldsymbol{\beta}$ from the (5.53);
   (b) Sample $\mathbf{u}$ from the (5.51);
   (c) Sample $\boldsymbol{\tau}^2$ from the (5.50);
   (d) Sample σ^2 from the (5.49).

The sampler is implemented in `C++` exploiting the `R` interface `Rcpp` (Eddelbuettel et al., 2011) and the matrix algebra library `RcppArmadillo` (Eddelbuettel and Sanderson, 2014). The `C` code that is used to generate random numbers from the GIG distribution is due to Gramacy (2010).
In the developed package `BayesLN` the function `LN_hierarchical` allows the user to specify the desired model through the simple specification of the model equation: the same syntax of the well known function `lemr` of the `lme4` package is adopted (Bates et al., 2015). Further details about the usage of this function will be provided in the subsequent sections.

5.3 Prior specification

In the previous sections, some plausible priors for the variance components have been listed, if the posterior moments of a log-normal mixed model are of interest. It is worth to highlight that the constraints on the hyperparameters might produce highly informative priors for the variances, since their tails are required to be notably light. In this sense, it is necessary to propose a weakly informative prior specification that avoids an excessive underestimation of the variance components and that preserves the balance among them. On the other hand, if any information about the problem is available, then it is possible to specify the triplet of hyperparameters considering the GIG properties like mean and variance, always taking into consideration the existence conditions.
In order to present the weakly informative strategy, the simplest case with two variance components σ^2 and τ^2 will be studied, then the results will be extended to the general situation. However, this is a remarkable case since it represents the random intercept model and in this framework it is possible to define the intraclass correlation coefficient:

$$\rho = \frac{\tau^2}{\sigma^2 + \tau^2}, \tag{5.82}$$

that is a quantity of interest in the analysis of hierarchical models, both from a statistical viewpoint and from the applied perspective.
By studying the marginal prior of this quantity when two GIG priors like (5.47) are specified for σ^2 and τ^2, some indications might be deduced in order to fix a weakly informative prior that respects the constraints on the parameters and, at the same time, it controls the prior balance among the variance components. In fact, the γ parameters could be fixed according to conditions of proposition 5.4 by evaluating the expressions with $(r+1)$, in order to assure the proper existence of the required moment of order r. Considering the three conditions:

(i) $\gamma_\sigma^2 = (r+1) + (r+1)^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}_i \right)$;

(ii) $\gamma_\sigma^2 = (r+1) + (r+1)^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}_i \right)$ and $\gamma_{\tau,s}^2 = (r+1) + (r+1)^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_{o,i}^T \mathbf{L}_s \tilde{\mathbf{x}}_{o_i} \right), \ \forall s$;

(iii) $\gamma_\sigma^2 = (r+1)^2 + (r+1)^2 \max_{i=1,\dots,\tilde{n}} \left(\tilde{\mathbf{x}}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tilde{\mathbf{x}}_i \right)$.

Several prior selections strategies pointed out that a uniform prior on the intraclass correlation coefficient leads to good frequentist properties for the parameters estimates. The idea of specifying a uniform distribution for ρ was presented by Chaloner (1987) and it constitutes the heuristic argument that justifies the *uniform shrinkage prior*, which have been extensively studied and used (Daniels, 1999; Natarajan and Kass, 2000).
As a first step, it is possible to observe that the prior on ρ follows a normalized generalized inverse Gaussian distribution: $\rho \sim N - GIG(\lambda_\tau, \delta_\tau, \gamma_\tau, \lambda_\sigma, \delta_\sigma, \gamma_\sigma)$, a distribution whose density has been derived and studied in a work by Favaro et al. (2012). The density is:

$$p(\rho) = \frac{\left(\frac{\gamma_\sigma}{\delta_\sigma}\right)^{\lambda_\sigma}\left(\frac{\gamma_\tau}{\delta_\tau}\right)^{\lambda_\tau}}{2K_{\lambda_\sigma}(\gamma_\sigma\delta_\sigma)K_{\lambda_\tau}(\gamma_\tau\delta_\tau)}\rho^{\lambda_\tau-1}(1-\rho)^{\lambda_\sigma-1}\left(\frac{\frac{\delta_\tau^2}{\rho}+\frac{\delta_\sigma^2}{1-\rho}}{\gamma_\tau^2\rho+\gamma_\sigma^2(1-\rho)}\right)^{\frac{\lambda_\tau+\lambda_\sigma}{2}} \times$$
$$\times K_{\lambda_\tau+\lambda_\sigma}\left(\sqrt{\left(\frac{\delta_\tau^2}{\rho}+\frac{\delta_\sigma^2}{1-\rho}\right)(\gamma_\tau^2\rho+\gamma_\sigma^2(1-\rho))}\right), \quad \rho \in (0,1). \tag{5.83}$$

To build a strategy that is based on the prior balance among variances, it is reasonable to assume the same marginal prior $GIG(\lambda, \delta, \gamma)$ for σ^2 and τ^2, inducing a prior on ρ controlled only by three hyperparameters, whose density is:

$$p(\rho) = \frac{K_{2\lambda}\left(\gamma^2\delta^2\left[\frac{1}{\rho}+\frac{1}{1-\rho}\right]\right)}{2\left[K_\lambda\left(\gamma\delta\right)\right]}\left[\rho(1-\rho)\right]^{\lambda-1}, \quad \rho \in (0,1). \tag{5.84}$$

Moreover, evaluating the target functionals of the analysis, the most restrictive threshold should be chosen as unique value of γ.
Because of the presence of ρ inside a Bessel K function, the analytic derivation of the distribution characteristics like moments is not possible. However, it is interesting to consider that the parameters δ and γ enter the density through a multiplication only. This fact allows to compensate high values of the tail parameter γ, caused by the constraints, by decreasing the value of the concentration parameter δ. A notable simplification of the prior on ρ is evident taking into consideration the limiting case $\delta \to 0$: it removes from $p(\rho)$ the potential effect of the constraint on γ and in previous applications of the GIG distribution as variance prior in the log-normal context this provides good frequentist properties (Fabrizi and Trivisano, 2012).
The density (5.84), in the limiting case $\delta \to 0$, becomes:

$$f(\rho) = \frac{\Gamma(|2\lambda|)}{\Gamma(|\lambda|)^2}\left[\rho(1-\rho)\right]^{|\lambda|-1}, \quad \rho \in (0,1); \tag{5.85}$$

exploiting the small argument approximation of the Bessel K function (A.7).
Remembering that the limiting case of the GIG distribution when $\lambda > 0$ and $\delta \to 0$ is the gamma distribution $Ga(\lambda, \beta)$, with $\beta = \gamma^2/2$ and having density:

$$f(x) = \frac{\beta^\lambda}{\Gamma(\lambda)}x^{\lambda-1}\exp\left\{-\beta x\right\}; \tag{5.86}$$

if the priors of the variances are gamma distributions with equal parameters, then the prior on ρ is a beta distribution with equal parameters λ. Moreover, if $\lambda = 1$ then $\rho \sim \mathcal{U}(0,1)$ a priori. In addition, if two different shape parameters λ_σ and λ_τ are chosen, then it is possible to control the prior information given on ρ.
Another quantity of interest in the variance components models is the ratio $\phi = \tau^2/\sigma^2$. It is a one-to-one transformation of ρ and strategies to fix its prior distribution have been studied (Chaloner, 1987; Ye, 1994). In particular, the latter proposed a solution for the problem in

the reference prior framework (Berger and Bernardo, 1992). Repeating the previous steps, then the prior of the ratio of variances GIG distributed has the following density:

$$\begin{aligned} p(\phi) &= \int_0^{+\infty} |\sigma^2| p_{\tau^2}\left(\phi\sigma^2\right) p_{\sigma^2}\left(\sigma^2\right) \mathrm{d}\sigma^2 \\ &= \frac{\left(\frac{\gamma_\tau}{\delta_\tau}\right)^{\lambda_\tau}\left(\frac{\gamma_\sigma}{\delta_\sigma}\right)^{\lambda_\sigma}}{2K_{\lambda_\sigma}(\gamma_\sigma\delta_\sigma)K_{\lambda_\tau}(\gamma_\tau\delta_\tau)}\phi^{\lambda_\tau-1}\left(\frac{\delta_\tau^2+\phi\delta_\sigma^2}{\phi}\frac{1}{\gamma_\tau^2\phi+\gamma_\sigma^2}\right)^{\frac{\lambda_\tau+\lambda_\sigma}{2}} \times \\ &\quad \times K_{\lambda_\tau+\lambda_\sigma}\left(\sqrt{\left(\delta_\tau^2+\phi\delta_\sigma^2\right)\left(\gamma_\tau^2+\gamma_\sigma^2/\phi\right)}\right), \end{aligned} \tag{5.87}$$

with equal hyperparameters. Then, letting $\delta \to 0$, the distribution $p(\phi)$ reduces to the following particular form of beta prime distribution:

$$f(\phi) = \frac{\Gamma(|2\lambda|)}{\Gamma(|\lambda|)^2}\phi^{|\lambda|-1}(1-\phi)^{-|2\lambda|}. \tag{5.88}$$

Finally, choosing $\lambda = 1$, the prior of the ratio is:

$$f(\phi) = (1+\phi)^{-2}. \tag{5.89}$$

The latter result confirms the prior for ϕ found by Chaloner (1987) when $\rho \sim \mathcal{U}(0,1)$ and is similar to the reference priors that was deduced by Ye (1994).

This strategy developed for models that have two variance components could be extended to the more general models with $q+1$ variances. The basic idea is to consider that for every $\rho_s = \tau_s(\tau_s+\sigma^2)^{-1}$ a uniform prior is specified. This occurs when all the priors are equal gamma distributions with shape 1 and scale parameter fixed respecting the existence condition.

To sum up, under the described setting, if the λ parameter is set to be positive, a gamma prior $\mathcal{G}\left(\lambda, \gamma^2/2\right)$ for each variance component is approximately assumed when $\delta = \varepsilon$. Clearly, γ^2 is considered fixed accordingly to the rules of theorem 5.3. As a consequence, a normal-gamma prior is specified marginally for the random effects vector $\mathbf{u}$. This prior setting is not new to the literature: it was introduced by Griffin and Brown (2010) as prior for the coefficients of a linear model. Its main characteristic is the notable degree of shrinkage towards 0 because of the light prior tail.

In the proposed setting with $\lambda = 1$, the gamma distribution degenerates to the exponential distribution, and in that case the normal-gamma is a Laplace distribution. This particular prior is known also as *Bayesian Lasso* and is characterized by a spike in 0. The degree of shrinkage determined by this prior could be further enhanced setting λ near to 0, whereas increasing the parameter value, the amount of shrinkage fixed decreases.

Fruhwirth-Schnatter and Wagner (2011) and Fabrizi et al. (2018) already used this prior also as prior for the random effects. The main difference between these applications and the present proposal is represented by the approach used to deal with the scale (or rate) parameter of the gamma prior. In fact, the cited works specify an hyperprior on it. Unfortunately, in this situation this solution is not viable because of the restrictions on the parameter space

due to the posterior moments existence condition. However, thanks to this interpretation of the proposed prior specification setting, an interpretation of the λ parameter is offered: its value controls the prior shrinkage and it could be fixed accordingly.
The `R` function `LN_hierarchical` automatically implements the prior strategy described in this section, choosing the more general existence condition after the user specifies the functionals of interest. In fact, the argument `functional` can receive as input the strings `"Marginal"` for $\theta_m(\tilde{\mathbf{x}})$, `"Subject"` for $\theta_c(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$ and `"PostPredictive"` for the posterior predictive distribution. The matrices $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Z}}$ can be provided as input of the respective arguments and, as point values for the tail parameters γ-s, the expression determining the existence conditions are evaluated with $r = 3$, in order to assure the well definition of the posterior mean and the posterior variance. The multiple points prediction setting is considered and the conditions of proposition 5.4 are implemented.

5.4 Applied focus: the small area estimation framework

The small area estimation (SAE) framework may represent an important field in which applying the methodologies developed in this work. For a generic introduction on the topic, see Pfeffermann et al. (2013) and Rao and Molina (2015).
If the area sample sizes are too narrow to apply the design based direct estimators, then the model-based approach has become largely diffused in SAE, in order to borrow strength among the areas and to fully exploit the whole disposable auxiliary information. Depending on the data structure, it is possible to distinguish between two main model building strategies: if the response is an areal aggregated quantity, then an *area level model* is considered; on the other hand, if the auxiliary information is available for each individual of the population, then a *unit level model* can be fitted. In this section, the second eventuality is examined.
Moreover, in this context, it is common to deal with positive and skewed response variables (e.g. business variables, areal surfaces, environmental variables) and the estimate of the area means is an even more delicate task.
In presence of skewed variables, it appears reasonable to assume log-normality for data. A synthetic estimator was derived under the log-normality assumption by Karlberg et al. (2000), whereas Chandra and Chambers (2009) deduced a model-based direct estimator.
Focusing on the model-based approach to SAE, mixed models are widely used since they allow to easily incorporate an area specific random effect. In the literature, it is possible to identify two main approaches to face their estimation: the first one is based on the empirical best linear unbiased prediction (EBLUP) estimation, that in the normal case coincides with the empirical Bayes (EB) approach. EBLUP and EB strategies to estimate area means under a log-normal mixed model have been developed by Berg and Chandra (2014), Molina and Martin (2018) and Zimmermann and Münnich (2018), who considered the informative sampling framework. The alternative estimation method is the hierarchical Bayes (HB) choice: it has a long tradition in SAE (Datta and Ghosh, 1991) and, among its strengths, it is possible to highlight the easiness of obtaining estimates of parameter transformations

naturally, deducing interval estimates, uncertainty measures and making predictions, taking advantage of the posterior predictive distribution.
However, recalling the findings of previous sections, it is worth to be careful in using HB with a log-normal model. For example, the solution proposed by Dagne (2001), who suggested a Box-Cox mixed model procedure, implies the non-finiteness of the posterior predictive distribution moments if the logarithmic transformation is found to be appropriate to normalize the response.
Thanks to the results of theorem 5.3, the aim of these sections is to fill the gap of the literature and propose a safe HB solution for analysing skewed variables in the SAE framework. After defining the notation (subsection 5.4.1), a brief review of the EB solution is hinted in section 5.4.2, whereas the HB method is exposed in section 5.4.3. Finally, in section 5.4.4, some considerations about the estimation of poverty measures are shown.

5.4.1 The unit level model

A finite population U constituted by N units is considered. It is partitioned into D sub-populations $U_1, ..., U_D$ having dimensions $N_1, ..., N_D$ such that $N = \sum_{d=1}^{D} N_d$. A random sample s with size n is drawn from the overall population U, obtaining D sub-samples $s_1, ..., s_D$ with sample sizes $n_1, ..., n_D$; $\sum_{d=1}^{D} n_d = n$. It is possible to denote with $\bar{s}_d$ the unsampled portion of the sub-population U_d that is of size $N_d - n_d$. The value assumed by the variable of interest for the unit $i \in \{1, ..., N_d\}$ belonging to area d, $d = 1, ..., D$ is denoted with y_{di} whereas the set of values observed for the p covariates is stored in the vector $\mathbf{x}_{di}$.
In this context, the usual inferential goals are the estimation of the amount of the variable of interest in an unsampled unit y_{di} in order to compute the area mean:

$$\bar{y}_d = N_d^{-1} \sum_{i=1}^{N_d} y_{di}. \tag{5.90}$$

Considering a positively skewed variable for which it is reasonable the log-normality assumption, then it is possible to define $w_{di} = \log y_{di}$ and specify the usual random intercept model by Battese et al. (1988) for the transformed variable. In particular:

$$\begin{aligned} &w_{di} = \mathbf{x}_{di}^T \boldsymbol{\beta} + u_d + e_{di}; \\ &u_d \sim \mathcal{N}(0, \tau^2),\ e_{di} \sim \mathcal{N}(0, \sigma^2);\ d = 1, ..., D;\ i = 1, ..., N_D. \end{aligned} \tag{5.91}$$

The model is a particular case of the log-normal linear mixed model faced in section 5.2, however other more general models like the ones included in Datta and Ghosh (1991) might be specified within the same framework.

5.4.2 Empirical Bayes estimation

In the EB estimation framework, the goal is to find the minimum MSE predictor of the area mean (5.90) under model (5.91). Exploiting the properties of the log-normal distribution it is possible to check that it coincides with:

$$\hat{\bar{y}}_d(\boldsymbol{\theta}) = N_d^{-1}\left[\sum_{i\in s_d} y_{di} + \sum_{i\in \bar{s}_d} \hat{y}_{di}(\boldsymbol{\theta})\right], \tag{5.92}$$

where the $\hat{y}_{di}(\boldsymbol{\theta})$ are the minimum MSE prediction of the unsampled units, which are function of the unknown parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2)$:

$$\hat{y}_{di}(\boldsymbol{\theta}) = \exp\left\{\mathbf{x}_{di}^T\boldsymbol{\beta} + \gamma_d\left(\bar{y}_{ds} - \bar{\mathbf{x}}_{ds}^T\boldsymbol{\beta}\right) + \frac{\sigma^2}{2}\left(\frac{\gamma_d}{n_d} - 1\right)\right\}. \tag{5.93}$$

For completeness, $\bar{y}_{ds}$ and $\bar{\mathbf{x}}_{ds}$ are the means of the response and of the covariates in area d, whereas $\gamma_d = \tau^2(\tau^2 + n_d^{-1}\sigma^2)^{-1}$ is the shrinkage factor.
Hence, the EB prediction is obtained by substituting the unknown parameters $\boldsymbol{\theta}$ with a consistent estimate $\hat{\boldsymbol{\theta}}$. In the subsequent sections of this work, the restricted maximum likelihood (REML) method is adopted, getting the final estimator $\hat{\bar{y}}_d(\hat{\boldsymbol{\theta}})$.
Another important point is the estimation of the MSE of $\hat{\bar{y}}_d(\hat{\boldsymbol{\theta}})$. It was derived analytically by Berg and Chandra (2014), but following Jiang et al. (2002) the jackknife method is used to compute it in the next sections. It is important to stress that the MSE of an EB predictor can be decomposed into two addends: a leading term, which is related to the variance of the minimum MSE predictor (5.92), and a second term that adjusts it for the uncertainty due to the estimation of the parameter vector $\boldsymbol{\theta}$.
Finally, an interval estimate proposal is the heuristic prediction interval used by Berg and Chandra (2014): $\hat{\bar{y}}_d(\hat{\boldsymbol{\theta}}) \pm 2.04\sqrt{\widehat{\text{MSE}}[\hat{\bar{y}}_d(\hat{\boldsymbol{\theta}})]}$.

5.4.3 Hierarchical Bayes estimation

In this section, the hierarchical Bayesian formulation of the popular random intercept model by Battese et al. (1988) is provided for the log-transformed response w_{di}. For each area d the following model in matrix form is specified:

$$\mathbf{w}_d = \mathbf{X}_d\boldsymbol{\beta} + \mathbf{1}_{N_d}u_d + \mathbf{e}_d,\ d = 1, ..., D; \tag{5.94}$$

where:

$$\begin{aligned}&\mathbf{w}_d = (w_{d1}, ..., w_{dN_d})^T \in \mathbb{R}^{N_d}, \quad \mathbf{e}_d = (e_{d1}, ..., e_{dN_d})^T \in \mathbb{R}^{N_d},\\ &\mathbf{X}_d = (\mathbf{x}_{d1}, ..., \mathbf{x}_{dN_d})^T \in \mathbb{R}^{N_d\times p}.\end{aligned} \tag{5.95}$$

Moreover, it is possible to partition the quantities related to area d by splitting them into sampled (with sd as subscript) and unsampled (with ud as subscript) units:

$$\mathbf{z}_d = (\mathbf{z}_{sd}^T\ \mathbf{z}_{ud}^T)^T, \quad \mathbf{e}_d = (\mathbf{e}_{sd}^T\ \mathbf{z}_{ud}^T)^T, \quad \mathbf{X}_d = (\mathbf{X}_{sd}^T\ \mathbf{X}_{ud}^T)^T. \tag{5.96}$$

As a consequence, the model parameters are estimated only using the n sampled units. Formally, the model equation is:

$$\mathbf{w}_s = \mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u} + \mathbf{e}_s, \tag{5.97}$$

where:

$$\begin{aligned}&\mathbf{w}_s = (\mathbf{w}_{s1}, ..., \mathbf{w}_{sD})^T \in \mathbb{R}^n, \ \mathbf{e}_s = (\mathbf{e}_{s1}, ..., \mathbf{e}_{sD})^T \in \mathbb{R}^n, \\ &\mathbf{X}_s = (\mathbf{X}_{s1}^T, ..., \mathbf{X}_{sD}^T)^T \in \mathbb{R}^{n\times p}, \quad \mathbf{Z}_s = \text{diag}_{d\in\{1,...,D\}} \mathbf{1}_{n_d} \in \mathbb{R}^{n\times D}.\end{aligned} \tag{5.98}$$

Hence, the complete formulation of the hierarchical model is:

$$\begin{aligned}&\mathbf{w}_s|\mathbf{u}, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n\left(\mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\mathbf{u}, \sigma^2\mathbf{I}_n\right); \\ &\mathbf{u}|\tau^2 \sim \mathcal{N}_D\left(0, \tau^2\mathbf{I}_D\right), \ (\boldsymbol{\beta}, \sigma^2) \sim p(\boldsymbol{\beta}, \sigma^2); \\ &\tau^2 \sim p(\tau^2).\end{aligned} \tag{5.99}$$

In this work, a flat improper prior is specified for the vector of coefficients $\boldsymbol{\beta}$, whereas prior independence is assumed for the variance components. In small area estimation applications, non-informative priors are often preferred in order to do not change too much the information carried out by data in topics like public policies or social studies by using subjective priors. For this reason, the prior specification method described in section 5.3 might be suitable for this task.

In order to properly fix the prior distributions for the variance components, it is worth to introduce the target quantity that is usually estimated in the SAE context. In fact, the aim of the estimation procedure is to preserve the existence of its posterior moments. As already hinted in the previous section, the inferential goal in SAE is the area mean $\bar{y}_d$. In the HB case, a natural way to estimate it is based on the use of the posterior predictive distribution. In fact, the vector of unsampled units of area d, $\mathbf{y}_{ud}$, can be estimated with the expectation of the posterior predictive distribution.

Considering the log-transformation $\mathbf{w}_{ud} = \log(\mathbf{y}_{ud})$, the predictive distribution conditioned on the parameters is:

$$\mathbf{w}_{ud}|\mathbf{u}, \boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_{N_d-n_d}\left(\mathbf{X}_{ud}\boldsymbol{\beta} + \mathbf{1}_{N_d-n_d}u_d, \sigma^2\mathbf{I}_{N_d-n_d}\right), \tag{5.100}$$

and the posterior predictive distribution $p(\mathbf{y}_{ud}|\mathbf{y}_s)$ (5.45) might be obtained through simulation methods. Then, it is natural to propose the following estimator for the vector on unsampled units of area d:

$$\hat{\mathbf{y}}_{ud}^{HB} = \mathbb{E}\left[\mathbf{y}_{ud}|\mathbf{y}_s\right]. \tag{5.101}$$

Recalling the result (iii) of theorem 5.3, the existence of the previous quantity can be only assured by a proper choice of the priors on the variance components.

Exploiting the MCMC methods, it is immediate to have an estimate of the area mean:

$$\hat{\bar{Y}}_d^{HB} = \frac{1}{N_d}\left(\sum_{i\in s_d} y_{di} + \sum_{i\in\bar{s}_d} \hat{y}_{di}^{HB}\right), \tag{5.102}$$

whose existence is assured by the finiteness of each addend.

Hence, assuming two equal GIG priors on the variance components, it is possible to fix their tail parameter fulfilling the condition:

$$\gamma^2 > r^2 + r^2 h_d^{max}, \quad h_d^{max} = \max_{i\in\bar{s}_d} \mathbf{x}_{di}^T(\mathbf{X}_s^T\mathbf{X}_s)\mathbf{x}_{di}. \tag{5.103}$$

More generally, if each area mean needs to be estimated, then the leverage to include in the condition is $h^{max} = \max_d h_d$. In practice, since the posterior variance is an inferential quantity of interest, r is usually set equal to 2.

5.4.4 The estimation of poverty measures

One of the main advantages of HB methods is that the estimates for non-linear functions of data can be easily provided. For example, it is common for SAE studies in economic and social fields to have the goal of investigating poverty. Among the plethora of poverty measures, the family due to Foster et al. (1984) (FGT) has been recently considered in the small area framework by Molina and Rao (2010) from an EB perspective and by Molina et al. (2014) using HB methods.
The family FGT of poverty measures is defined for subject i of area d as:

$$F_{di,\alpha} = \left(\frac{c - y_{di}}{c}\right)^{\alpha} \mathbf{1}_{\{y_{di}<c\}}(Y_{di}); \tag{5.104}$$

where y_{di}, in this case, is the welfare variable considered in the study (e.g. income or expenditure) and c is the poverty line, i.e. the value of the response below which a subject is labelled as *experiencing poverty*. By changing α, it is possible to express different poverty measures: when $\alpha = 0$ the expression reduces to the indicator function and it is called *poverty incidence*, if $\alpha = 1$ the *poverty gap* is obtained and can be viewed as the distance of an individual under poverty to the poverty line, and finally, with $\alpha = 2$ larger deviations are emphasized and the *poverty severity* measure is computed.
Therefore, the inferential problem involves the estimation of a non-linear function of the response variable under study. Moreover, as it is well known, many economic variables are positive and skewed: in fact, the two works cited before specify the classical Battese, Harter and Fuller model for a transformation of the studied response. Even if they developed a theoretical framework considering a general transformation, in practice they always use the logarithm to transform data, fitting the random intercept model with $\mathbf{w}_d = \log \mathbf{y}_d$ as a response, like model (5.91).
In this case, the area mean of the poverty measure is the target quantity under study:

$$\bar{F}_{d,\alpha} = \frac{1}{N_d} \sum_{i=1}^{N_d} F_{di,\alpha}, \quad d = 1, .., D. \tag{5.105}$$

Exploiting the model-based inference and the partition of the population in sampled or unsampled units, in the EB estimation procedure it is possible to consider a minimum MSE estimator parallel to the (5.92):

$$F_{d,\alpha} = \frac{1}{N_d} \left[\sum_{i \in s_d} F_{di,\alpha} + \sum_{i \in \bar{s}_d} \hat{F}^{EB}_{di,\alpha} \right], \tag{5.106}$$

and the single predictions of the subjects poverty measures are computed using the conditioned predictive density (5.100):

$$\begin{aligned}\hat{F}^{EB}_{dj,\alpha} &= \mathbb{E}\left[F_{dj,\alpha}|\mathbf{y}_s\right] \\ &= \int_{-\infty}^{+\infty} \left(\frac{c-\exp(w_{di})}{c}\right)^{\alpha} \mathbf{1}_{\{\exp(w_{di})<c\}}(w_{di}) p_{\mathcal{N}}\left(w_{di};\mu_{di|s},V_{di|s}\right) \mathrm{d}w_{di},\end{aligned} \tag{5.107}$$

where $\mu_{di|s}$ and $V_{di|s}$ are the mean and the variance that need to be estimated for the unsampled unit w_{di} through the available sample.
According to the authors, these integrals cannot be analytically solved because of their complex structure: they used MC methods with a consequent computationally expensive inferential procedure. However, when the logarithm transformation is adopted, it is possible to express these quantities through cumulative functions of the normal distribution $\Phi(\cdot)$ with a huge gain in terms of computing time.
Returning to the main topic of this work, it is worth to focus on the HB estimation of poverty measures analysing the proposal by Molina et al. (2014). As before, the expectation of the predictive posterior distribution is chosen to estimate $F_{di,\alpha}$ for the unsampled units. Under the model (5.99) it is:

$$\begin{aligned}\hat{F}^{HB}_{di,\alpha} &= \mathbb{E}\left[F_{di,\alpha}|\mathbf{y}_s\right] \\ &= \int_{\Theta}\int_{-\infty}^{+\infty} \left(\frac{c-\exp(w_{di})}{c}\right)^{\alpha} \mathbf{1}_{\{\exp(w_{di})<c\}}(w_{di}) p_{\mathcal{N}}\left(w_{di};\mathbf{x}_j^T\boldsymbol{\beta}+u_d,\sigma^2\right) \times \\ &\qquad p(u_d,\boldsymbol{\beta},\sigma^2,\tau^2|\mathbf{y}_s)\mathrm{d}w_{di}\mathrm{d}\boldsymbol{\theta}.\end{aligned} \tag{5.108}$$

With simple computations, it can be proved that the existence conditions of the latter integral coincides with the point (iii) of theorem 5.3, provided that $r=\alpha$. Moreover, if the generic moment of order r' of the posterior predictive distribution is required, then $r=\alpha\cdot r'$. As a consequence, no issue of existence involves the poverty incidence ($\alpha=0$), as might be expected, whereas particular attention must be paid in specifying the priors when $\alpha=1,2$. In their work, Molina et al. (2014) fixed the improper uniform shrinkage prior $p(\sigma^2,\tau^2)\propto(\sigma^2+\tau^2)^{-2}$, which coincides to the (2.5) in Chaloner (1987), and that clearly leads to a posterior predictive distribution with infinite moments, according to proposition 5.3. Hence, the prior specification strategy described in section 5.3 should be followed to carry out a proper inference.

Chapter 6

Log-normal mixed model: simulations and examples

In this chapter, the theoretical findings about the Bayesian estimation of the log-normal mixed model are applied: firstly, a simulation study to assess the frequentist properties of the proposed estimators is reported (section 6.1), finally the methodology is used in real data examples (section 6.2).

6.1 Simulation study

The simulation study is planned to investigate two different applied aspects. In the first place, to assess the frequentist properties of the estimates of the population means in the linear mixed model framework, the simple one-way ANOVA random effect model is considered (section 6.1.1). Then, the focus moved to the behaviour of the developed methodology in small area estimation problems: both a model-based simulation and a design-based simulation on the classical BHF model are carried out (section 6.1.2).

6.1.1 ANOVA model

The balanced one-way random effect ANOVA is chosen as generating model for data in the simulation study. It is the particular case of the model studied in section 5.1, in which all groups has sample size n_g:

$$w_{ij} = \log(y_{ij}) = \mu + \nu_j + \varepsilon_{ij};\ j = 1, ..., m;\ i = 1, ..., n_g; \tag{6.1}$$

where the random effects $\nu_j \sim \mathcal{N}\left(0, \tau^2\right)$ are independent with respect to the unstructured error terms $\varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right)$.
The inferential task consists of estimating the marginal expectation θ_m, reported in equation (5.4) and expectation conditioned on the random effects $\theta_c(\nu_j)$, in equation (5.5), for each group. Besides, the estimators derived in chapter 5 following the fully Bayesian methodology

with priors specified as indicated in section 5.3:

$$p(\mu) \propto 1, \quad \sigma^2 \sim GIG(1, 0.01, \gamma_{\text{m}}), \ \tau^2 \sim GIG(1, 0.01, \gamma_{\text{m}}), \tag{6.2}$$

are denoted as $\hat{\theta}_m^{B,GIG}$ and $\hat{\theta}_c^{B,GIG}(\nu_j)$, where the notation $\gamma_{\text{m}} = \max\{\gamma_\sigma, \gamma_{\tau,s}\}$. The Bayes estimators under the improper uniform shrinkage prior (Chaloner, 1987):

$$p(\mu) \propto 1, \quad p(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \rho \sim \mathcal{U}(0, 1), \tag{6.3}$$

are labelled as $\hat{\theta}_m^{B,US}$ and $\hat{\theta}_c^{B,US}(\nu_j)$. Finally, the proper half-t priors for scales suggested by Gelman (2006):

$$\mu \sim \mathcal{N}(0, 100), \quad \sigma \sim \text{half-}t_3, \quad \tau \sim \text{half-}t_3, \tag{6.4}$$

produces the estimators $\hat{\theta}_m^{B,t}$ and $\hat{\theta}_c^{B,t}(\nu_j)$ that are included in the study too. The algorithm for computing the posteriors in the uniform shrinkage prior is adapted from the one in Molina et al. (2014), whereas the model with half-t priors with three degrees of freedom are implemented in `Stan` (Carpenter et al., 2017) using the interface package `brms` (Bürkner, 2018). In addition, the frequentist estimates obtained with the classical function `lmer` with REML as fitting method are evaluated. Concerning the global expectation, a plug-in estimator of θ_m is examined:

$$\hat{\theta}_m^{REML} = \exp\left\{\hat{\mu}^{REML} + \frac{\hat{\sigma}^2_{REML} + \hat{\tau}^2_{REML}}{2}\right\}. \tag{6.5}$$

Then, considering the expectation conditioned on the random effects, the particular case of the empirical Bayes estimator $\hat{\theta}_c^{EB}(\nu_j)$ described in section 5.4.2 is computed. It has been proposed for this model also by Rappaport et al. (1995). Finally, the properties of Bayes estimator under relative quadratic loss conditioned on the variance components (section 5.1.1) are evaluated in order to provide a benchmark for the other estimators.

A total amount of $B = 2000$ samples are generated from the one-way random effect ANOVA model and a total of 36 scenarios are obtained crossing the following parameters choices: $n_g = (2, 5)$, $m = (10, 20)$, $\phi = \tau^2/\sigma^2 = (0.5, 1, 2)$ and $\sigma^2 = (0.5, 1, 2)$. The overall mean in the logarithmic scale is kept fixed at $\mu = 0$. The estimates that require Monte Carlo methods are based on 4000 iterations and the first 1000 iterations are discarded as burn-in. The convergence of the MCMC algorithm has been checked on a sample of replicates and no issues were observed.

Both the frequentist properties of the point estimator and of the interval estimates, for the methodologies in which they are available, are monitored. Bias, root mean square error (RMSE), frequentist coverage and average interval width are reported for the generic estimator $\hat{\theta}_m$ of the marginal expectation:

$$\begin{aligned} Bias(\hat{\theta}_m) &= \frac{1}{B}\sum_{k=1}^{B}\left(\hat{\theta}_m^{(k)} - \theta_m\right), \\ RRMSE(\hat{\theta}_m) &= \sqrt{\frac{1}{B}\sum_{k=1}^{B}\left(\hat{\theta}_m^{(k)} - \theta_m\right)^2}, \\ Cov(\hat{\theta}_m) &= \frac{1}{B}\sum_{k=1}^{B}\mathbf{1}_{\left[\hat{L}^{(k)};\hat{U}^{(k)}\right]}(\theta_m), \\ Wid(\hat{\theta}_m) &= \frac{1}{B}\sum_{k=1}^{B}\left(\hat{U}^{(k)} - \hat{L}^{(k)}\right); \end{aligned} \tag{6.6}$$

where $\hat{\theta}_m^{(k)}$ is the estimate of the true overall expectation θ_m at Monte Carlo iteration k and $\hat{L}^{(k)}$ and $\hat{U}^{(k)}$ are the estimated lower bound and upper bound for the 95% intervals.
On the other hand, to jointly evaluate the m different estimates for $\theta_c(\nu_j),\ j = 1, ..., m$, an average evaluation of the estimates $\hat{\hat{\theta}}_c$ is required. Therefore, the relative absolute bias (RABias), the relative RMSE (RRMSE), the average frequentist coverage (ACo.) and the average interval width (AWi.) are studied:

$$\begin{aligned} RABias(\hat{\hat{\theta}}_c) &= \frac{1}{J}\sum_{j=1}^{J}\left|\frac{1}{B}\sum_{k=1}^{B}\left(\frac{\hat{\theta}_c^{(k)}(\nu_j) - \theta_c^{(k)}(\nu_j)}{\theta_c^{(k)}(\nu_j)}\right)\right|, \\ RRMSE(\hat{\hat{\theta}}_c) &= \frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{1}{B}\sum_{k=1}^{B}\left(\frac{\hat{\theta}_c^{(k)}(\nu_j) - \theta_c^{(k)}(\nu_j)}{\theta_c^{(k)}(\nu_j)}\right)^2}, \\ ACo(\hat{\hat{\theta}}_c) &= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{B}\sum_{k=1}^{B}\mathbf{1}_{\left[\hat{L}^{(k)}(\nu_j);\hat{U}^{(k)}(\nu_j)\right]}\left(\theta_c^{(k)}(\nu_j)\right), \\ AWi(\hat{\hat{\theta}}_c) &= \frac{1}{J}\sum_{j=1}^{J}\frac{1}{B}\sum_{k=1}^{B}\left(\hat{U}^{(k)}(\nu_j) - \hat{L}^{(k)}(\nu_j)\right); \end{aligned} \tag{6.7}$$

where $\hat{\theta}_c^{(k)}(\nu_j)$ is the estimate of the j-th true group specific expectation $\theta_c^{(k)}(\nu_j)$ at Monte Carlo iteration k; $\hat{L}^{(k)}(\nu_j)$ and $\hat{U}^{(k)}(\nu_j)$ are the estimated lower bound and upper bound for the 95% intervals.

						Cond. Bayes		REML		U.S. prior		Half-t prior		GIG prior	
n_g	m	ϕ	σ^2	Scen.	θ_m	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
			0.5	1	1.45	-0.07	0.32	0.05	**0.41**	0.54	9.70	0.37	1.55	0.24	0.52
		0.5	1	2	2.12	-0.20	0.65	0.20	**1.06**	$> 10^5$	$> 10^6$	$> 10^3$	$> 10^5$	0.37	1.07
			2	3	4.48	-0.82	1.89	1.34	5.23	$> 10^{16}$	$> 10^{18}$	$> 10^5$	$> 10^6$	0.11	**2.82**
			0.5	4	1.65	-0.12	0.45	0.10	**0.63**	202.51	$> 10^3$	0.88	5.80	0.28	0.71
	10	1	1	5	2.72	-0.38	1.02	0.47	2.05	$> 10^{10}$	$> 10^{12}$	$> 10^9$	$> 10^{10}$	0.35	**1.60**
			2	6	7.39	-1.91	3.76	4.63	16.92	$> 10^{28}$	$> 10^{29}$	$> 10^9$	$> 10^{11}$	-0.98	**5.16**
			0.5	7	2.12	-0.24	0.74	0.27	1.29	$> 10^4$	$> 10^6$	830.70	$> 10^4$	0.32	**1.15**
		2	1	8	4.48	-0.98	2.13	1.87	6.99	$> 10^{16}$	$> 10^{18}$	$> 10^7$	$> 10^9$	-0.06	**3.08**
2			2	9	20.09	-7.82	12.66	44.93	188.79	$> 10^{40}$	$> 10^{41}$	$> 10^{25}$	$> 10^{26}$	-8.61	**15.46**
			0.5	10	1.45	-0.04	0.23	0.02	**0.27**	0.11	0.32	0.12	0.33	0.12	0.32
		0.5	1	11	2.12	-0.11	0.46	0.09	**0.66**	0.41	0.94	0.40	0.91	0.23	0.72
			2	12	4.48	-0.46	1.37	0.56	2.82	3.92	22.18	2.40	6.13	0.22	**2.12**
			0.5	13	1.65	-0.07	0.32	0.04	**0.40**	0.18	0.49	0.21	0.51	0.16	0.45
	20	1	1	14	2.72	-0.21	0.73	0.19	1.19	0.98	5.16	0.80	1.82	0.25	**1.11**
			2	15	7.39	-1.08	2.75	1.77	7.63	$> 10^4$	$> 10^6$	12.55	67.75	-0.45	**3.95**
			0.5	16	2.12	-0.14	0.52	0.11	**0.76**	0.41	1.04	0.47	1.08	0.20	**0.76**
		2	1	17	4.48	-0.55	1.54	0.71	3.35	5.86	41.50	3.68	11.26	0.03	**2.21**
			2	18	20.09	-4.57	9.45	13.16	51.99	$> 10^6$	$> 10^8$	$> 10^5$	$> 10^7$	-6.44	**12.39**
			0.5	19	1.45	-0.05	0.27	0.03	**0.32**	0.20	1.22	0.20	0.45	0.16	0.39
		0.5	1	20	2.12	-0.15	0.55	0.10	**0.78**	730.67	$> 10^4$	1.16	20.26	0.27	0.83
			2	21	4.48	-0.60	1.62	0.67	3.43	$> 10^{11}$	$> 10^{13}$	$> 10^4$	$> 10^5$	0.23	**2.39**
			0.5	22	1.65	-0.10	0.40	0.06	**0.52**	$> 10^4$	$> 10^6$	0.57	4.06	0.21	0.59
	10	1	1	23	2.72	-0.31	0.93	0.29	1.61	$> 10^{16}$	$> 10^{18}$	142.19	$> 10^3$	0.29	**1.36**
			2	24	7.39	-1.59	3.44	2.81	12.36	$> 10^{40}$	$> 10^{42}$	$> 10^7$	$> 10^8$	-0.58	**4.69**
			0.5	25	2.12	-0.22	0.70	0.20	1.12	$> 10^4$	$> 10^6$	5.87	85.07	0.25	**1.01**
		2	1	26	4.48	-0.88	2.02	1.37	5.78	$> 10^{16}$	$> 10^{18}$	$> 10^7$	$> 10^9$	-0.05	**2.78**
5			2	27	20.09	-7.09	12.11	31.80	158.18	$> 10^{40}$	$> 10^{42}$	$> 10^{17}$	$> 10^{18}$	-7.53	**14.57**
			0.5	28	1.45	-0.03	0.20	0.01	**0.22**	0.07	0.25	0.08	0.25	0.08	0.25
		0.5	1	29	2.12	-0.08	0.40	0.05	**0.51**	0.24	0.63	0.23	0.63	0.17	0.56
			2	30	4.48	-0.31	1.18	0.29	1.90	1.49	3.17	1.23	4.04	0.27	**1.73**
			0.5	31	1.65	-0.05	0.29	0.03	**0.35**	0.13	0.40	0.15	0.42	0.12	0.38
	20	1	1	32	2.72	-0.16	0.67	0.12	**0.95**	0.56	1.34	0.55	1.32	0.21	**0.95**
			2	33	7.39	-0.85	2.53	1.06	5.21	8.51	40.11	5.21	13.45	-0.14	**3.52**
			0.5	34	2.12	-0.12	0.50	0.08	**0.68**	0.32	0.87	0.37	0.92	0.16	0.69
		2	1	35	4.48	-0.48	1.47	0.51	2.74	2.76	7.13	2.56	6.93	0.04	**2.03**
			2	36	20.09	-3.99	9.07	8.70	34.20	$> 10^4$	$> 10^5$	$> 10^3$	$> 10^4$	-5.36	**11.51**

Table 6.1: Bias and RMSE for the considered estimators of θ_m in the different scenarios.

					Cond. Bayes		REML		U.S. prior		Half-t prior		GIG prior	
n_g	m	ϕ	σ^2	Scen.	RABias	RRMSE	RABias	RRMSE	RABias	RRMSE	RABias	RRMSE	RABias	RRMSE
			0.5	1	0.13	0.36	0.13	**0.52**	0.23	0.57	0.25	0.59	0.20	0.53
		0.5	1	2	0.24	0.49	0.28	0.93	0.61	1.27	0.59	1.26	0.34	**0.89**
			2	3	0.43	0.65	0.74	2.24	5.82	64.48	2.97	17.02	0.52	**1.59**
			0.5	4	0.16	0.40	0.17	**0.61**	0.35	0.75	0.35	0.78	0.27	0.67
	10	1	1	5	0.30	0.54	0.39	**1.16**	1.02	2.10	1.07	6.04	0.51	1.25
			2	6	0.51	0.70	1.07	3.20	79.33	$> 10^3$	48.65	$> 10^3$	0.95	**2.81**
			0.5	7	0.19	0.43	0.20	**0.65**	0.53	1.03	0.48	1.05	0.36	0.84
		2	1	8	0.34	0.57	0.47	**1.30**	2.62	26.42	1.75	6.03	0.74	1.82
2			2	9	0.56	0.74	1.35	**3.89**	$> 10^6$	$> 10^7$	$> 10^4$	$> 10^6$	1.68	5.90
			0.5	10	0.12	0.35	0.13	**0.49**	0.18	0.50	0.19	0.52	0.17	0.50
		0.5	1	11	0.23	0.48	0.28	**0.86**	0.41	0.95	0.44	1.00	0.34	0.87
			2	12	0.41	0.63	0.68	1.87	1.17	2.50	1.21	2.70	0.66	**1.69**
			0.5	13	0.16	0.39	0.16	**0.57**	0.26	0.64	0.27	0.67	0.24	0.63
	20	1	1	14	0.29	0.53	0.37	**1.05**	0.64	1.36	0.65	1.44	0.51	1.19
			2	15	0.50	0.70	0.95	**2.59**	2.11	4.74	2.20	5.71	1.10	2.82
			0.5	16	0.18	0.43	0.19	**0.62**	0.36	0.79	0.33	0.78	0.31	0.74
		2	1	17	0.33	0.57	0.44	**1.18**	0.94	1.89	0.85	1.93	0.68	1.56
			2	18	0.56	0.74	1.17	**3.08**	4.00	13.14	4.06	19.81	1.67	4.71
			0.5	19	0.07	0.27	0.09	0.35	0.11	0.35	0.11	0.35	0.10	**0.33**
		0.5	1	20	0.14	0.37	0.19	0.58	0.26	0.61	0.23	0.60	0.18	**0.54**
			2	21	0.26	0.50	0.45	1.13	0.66	1.33	0.56	1.26	0.30	**0.91**
			0.5	22	0.08	0.29	0.11	0.38	0.15	0.39	0.12	0.37	0.11	**0.36**
	10	1	1	23	0.16	0.39	0.24	0.65	0.34	0.71	0.26	0.65	0.21	**0.60**
			2	24	0.29	0.53	0.57	1.33	0.89	1.69	0.64	1.44	0.38	**1.07**
			0.5	25	0.09	0.30	0.12	0.39	0.19	0.43	0.12	0.38	0.12	**0.38**
		2	1	26	0.17	0.41	0.27	0.69	0.45	0.84	0.27	0.68	0.23	**0.64**
5			2	27	0.31	0.55	0.66	1.46	1.25	2.20	0.68	1.52	0.43	**1.18**
			0.5	28	0.07	0.27	0.09	0.34	0.10	0.33	0.09	0.33	0.09	**0.32**
		0.5	1	29	0.14	0.37	0.20	0.55	0.21	0.54	0.19	0.54	0.18	**0.52**
			2	30	0.25	0.50	0.44	1.05	0.48	1.05	0.45	1.04	0.34	**0.91**
			0.5	31	0.08	0.28	0.11	0.37	0.12	0.36	0.10	0.35	0.10	**0.35**
	20	1	1	32	0.15	0.39	0.24	0.62	0.26	0.61	0.22	0.59	0.20	**0.57**
			2	33	0.29	0.53	0.56	1.24	0.61	1.26	0.51	1.15	0.41	**1.04**
			0.5	34	0.09	0.29	0.13	0.38	0.14	0.39	0.11	0.37	0.11	**0.37**
		2	1	35	0.17	0.41	0.27	0.66	0.31	0.68	0.23	0.62	0.22	**0.61**
			2	36	0.31	0.55	0.64	1.36	0.76	1.47	0.54	1.23	0.45	**1.13**

Table 6.2: RABias and RRMSE for the considered estimators of the group-specific expectations in the different scenarios.

						Global Predictor						Conditioned Predictor					
						U.S. prior		Half-t prior		GIG prior		U.S. prior		Half-t prior		GIG prior	
n_g	m	ϕ	σ^2	Scen.	θ_m	Cov.	Wid.	Cov.	Wid.	Cov.	Wid.	ACo	AWi	ACo	AWi	ACo	AWi
			0.5	1	1.45	0.95	2.59	0.94	2.71	0.95	2.19	0.97	3.20	0.94	2.75	0.95	2.62
		0.5	1	2	2.12	0.95	10.38	0.95	8.85	0.95	4.42	0.97	8.66	0.94	6.92	0.93	5.20
			2	3	4.48	0.95	186.11	0.95	72.18	0.90	11.01	0.97	51.29	0.93	33.60	0.90	13.20
			0.5	4	1.65	0.94	4.27	0.94	4.52	0.94	2.90	0.97	4.36	0.94	3.60	0.94	3.23
	10	1	1	5	2.72	0.95	27.21	0.95	21.31	0.94	6.27	0.97	14.23	0.93	10.47	0.92	7.10
			2	6	7.39	0.95	$> 10^3$	0.94	485.62	0.82	17.50	0.97	127.80	0.93	71.85	0.89	21.70
			0.5	7	2.12	0.94	11.30	0.94	11.72	0.93	4.46	0.98	7.20	0.94	5.31	0.94	4.49
		2	1	8	4.48	0.94	236.77	0.94	133.21	0.86	10.89	0.98	33.70	0.94	20.42	0.92	12.07
2			2	9	20.09	0.95	$> 10^5$	0.93	$> 10^4$	0.58	37.09	0.98	677.93	0.93	258.62	0.89	55.06
			0.5	10	1.45	0.96	1.32	0.96	1.33	0.96	1.32	0.95	2.49	0.93	2.26	0.94	2.32
		0.5	1	11	2.12	0.95	3.72	0.95	3.59	0.96	2.95	0.95	5.81	0.93	5.11	0.93	4.82
			2	12	4.48	0.95	22.43	0.95	18.00	0.94	8.79	0.95	23.19	0.92	18.87	0.92	13.53
			0.5	13	1.65	0.95	1.99	0.95	2.05	0.95	1.85	0.96	3.34	0.93	3.00	0.94	2.96
	20	1	1	14	2.72	0.95	7.24	0.95	6.91	0.94	4.53	0.96	9.00	0.93	7.73	0.93	6.83
			2	15	7.39	0.95	82.10	0.95	58.88	0.89	15.49	0.96	48.52	0.93	37.17	0.92	23.22
			0.5	16	2.12	0.95	4.10	0.95	4.28	0.94	3.11	0.97	5.12	0.94	4.35	0.94	4.16
		2	1	17	4.48	0.95	26.48	0.95	24.18	0.91	8.85	0.97	18.33	0.94	14.38	0.93	11.84
			2	18	20.09	0.95	$> 10^3$	0.95	815.10	0.73	38.11	0.97	175.11	0.94	113.55	0.92	59.64
			0.5	19	1.45	0.95	1.76	0.95	1.85	0.95	1.64	0.98	2.16	0.94	1.71	0.94	1.71
		0.5	1	20	2.12	0.95	5.56	0.95	5.26	0.95	3.42	0.98	4.99	0.93	3.80	0.94	3.52
			2	21	4.48	0.96	48.69	0.95	31.69	0.92	9.40	0.98	19.46	0.93	13.41	0.92	9.88
			0.5	22	1.65	0.94	3.13	0.94	3.47	0.95	2.41	0.98	2.90	0.94	2.09	0.94	2.04
	10	1	1	23	2.72	0.95	15.32	0.94	14.15	0.94	5.38	0.98	7.82	0.94	5.25	0.93	4.70
			2	24	7.39	0.95	527.82	0.94	231.62	0.85	16.54	0.98	42.08	0.94	23.77	0.92	16.40
			0.5	25	2.12	0.94	8.59	0.94	9.60	0.93	4.00	0.99	4.66	0.95	2.80	0.94	2.69
		2	1	26	4.48	0.94	135.65	0.94	98.70	0.87	10.10	0.99	17.29	0.94	9.03	0.93	7.81
5			2	27	20.09	0.94	$> 10^5$	0.94	$> 10^4$	0.64	37.95	0.99	177.29	0.94	68.32	0.92	43.03
			0.5	28	1.45	0.95	0.99	0.95	1.01	0.95	1.01	0.96	1.81	0.94	1.62	0.94	1.63
		0.5	1	29	2.12	0.95	2.55	0.95	2.50	0.95	2.25	0.96	3.96	0.94	3.46	0.94	3.39
			2	30	4.48	0.95	11.94	0.95	10.60	0.93	7.10	0.96	13.50	0.94	11.28	0.93	10.01
			0.5	31	1.65	0.95	1.62	0.94	1.69	0.95	1.56	0.97	2.35	0.94	1.98	0.95	1.97
	20	1	1	32	2.72	0.95	5.32	0.95	5.26	0.94	3.85	0.97	5.87	0.94	4.80	0.94	4.60
			2	33	7.39	0.95	44.17	0.94	36.91	0.90	14.08	0.97	26.34	0.94	20.06	0.93	16.97
			0.5	34	2.12	0.94	3.50	0.94	3.71	0.94	2.81	0.98	3.49	0.95	2.64	0.95	2.61
		2	1	35	4.48	0.94	19.69	0.94	18.96	0.91	8.17	0.98	11.46	0.95	8.19	0.94	7.68
			2	36	20.09	0.95	665.01	0.94	474.91	0.77	38.12	0.98	88.99	0.95	56.56	0.93	44.55

Table 6.3: Coverage and average width for the credible intervals of θ_m and averaged for the group-specific expectations.

The results reported in table 6.1 concern the performances of the marginal expectation point estimators. For each scenario, the true values θ_m are included in the table. Evident issues with the priors that do not assure the posterior moments existence for the target functionals can be individuated. In particular, in case of small sample sizes or moderate values of the variance terms, both the bias and the RMSE of $\hat{\theta}_m^{B,US}$ and $\hat{\theta}_m^{B,t}$ result excessively elevated and completely inappropriate to the scale of the inferential task. On the other hand, in scenarios with moderate sample size and very low values of variance components, the issues related to the non-finiteness of moments are masked and inference might be unconsciously carried out, even if it would be meaningless form a mathematical point of view.
As expected, the conditioned Bayes estimator under relative quadratic loss always assumes the minimum RMSE value and it is distinguished by negative bias coherently with previous findings in literature (Fabrizi and Trivisano, 2012). Among the unconditioned estimators, with the exception of scenarios characterized by small variances in which the $\hat{\theta}_m^{REML}$ has the lower RMSE, $\hat{\theta}_m^{B,GIG}$ is the more efficient estimator. In particular, it is the only choice having RMSE comparable to the conditioned estimator in extreme cases with large variance terms: this result is attained enhancing a negative bias. On the other hand, this bias has an unfavourable consequence on the frequentist coverage of the credible intervals reported in table 6.3. The coverage is largely below the fixed level of 0.95 in scenarios with $\sigma^2 = 2$. However, with the same conditions, the other prior settings considered produce intervals that reach the nominal level at the price of an average width excessively high. Therefore, in that population settings, which are particularly extreme in the log-normal framework, all the proposed intervals have little utility in practice even if for different reasons. In less extreme scenarios, in which the credible intervals with GIG priors attain the nominal coverage, the average width is always lower than the widths of intervals with other priors, hence they are more precise.
In table 6.2 the averaged results about the group-specific expectations estimates are reported. Also in this case, the lowest RRMSEs are observed for the conditioned Bayes estimator under relative quadratic loss. The naive Bayesian posterior means $\hat{\theta}_c^{B,US}(\nu_j)$ and $\hat{\theta}_c^{B,t}(\nu_j)$ for all j are always overcome in terms of RRMSE by $\hat{\theta}_c^{B,GIG}(\nu_j)$, even if they rarely show excessive values due to the non-finite moments of the posterior, as it was evident in the case of θ_m. Comparing the RRMSE of $\hat{\theta}_c^{B,GIG}(\nu_j)$ to the one of the empirical Bayes estimators $\hat{\theta}_c^{EB}(\nu_j)$, two different conditions can be distinguished. When $n_g = 2$, the EB estimator tends to show a lower RRMSE than the Bayes estimator with GIG priors, whereas if the group size increases ($n_g = 5$), $\hat{\theta}_c^{B,GIG}(\nu_j)$ is characterized by systematically better frequentist properties. The first feature might be caused by an underestimation of the variance component τ^2 due to the uniform prior on the intraclass correlation coefficient and a little sample information available ($n_g = 2$). The average coverage of the credible intervals under GIG prior underestimates the nominal level in the same extreme scenarios of θ_m, even if the departure from the target value 0.95 is slighter. However, when the nominal coverage level is reached, they are always shorter and, consequently, more powerful. Finally, it must be highlighted that the uniform shrinkage prior tends to produce intervals too wide with a consequent frequentist over-coverage.

Prior specification tuning

It is worth to emphasize that the GIG prior proposed in section 5.3 and analysed in this simulation exercise is considered in a generic formulation: if some prior information is available for the user, then some adjustments might be operated. For example, if a particularly high variance component in the logarithmic scale is expected, then a slight change in the prior scale can be considered: the default proposal is to fix it very small through the setting $\delta = \varepsilon$. This choice is generally good for typical log-normal scenarios, but in different conditions the prior scale could be increased to avoid the possible underestimation of the variance component (e.g. $\delta = 1$).
Besides, as already hinted in section 5.3, if a prior knowledge about the unbalance among variance component suggests a non-uniform prior on the intraclass correlation coefficient, then the shape parameters λ can be set recalling that they act as the parameters of a beta prior on ρ.
Finally, recalling the considerations about the normal-gamma priors on the random effects, the value of λ could be decided according to the desired amount of shrinkage. For example, the case $\lambda = 0.5$ (higher shrinkage) and $\lambda = 2$ (lower shrinkage) are evaluated. In fact, a further simulation study is carried out in order to evaluate the estimates obtained with these prior in the 36 scenarios. From the results reported in tables D.3 and D.4 in the appendix, the choice $\lambda = 0.5$ represents a reasonable strategy if there is a belief that $\sigma^2 < 0.5$. In fact, in that case it produces point estimates for θ_m which reach the REML performances, but the frequentist coverage of the credible intervals rapidly deteriorate as σ^2 increases. On the other hand, with $\lambda = 2$, interesting results are obtained: the properties of the point estimators appear to be slightly worst, but the frequentist coverages considerably improve in the extreme scenarios.

6.1.2 SAE framework

The second part of the simulation study deals with the frequentist properties of the log-normal linear mixed model estimates in the small area estimation framework described in section 5.4. Firstly, a model-based simulation study based on the unit-level random intercept model with a log-normal response and a single covariate is performed. The parameters of the model and the dimensions of population and sample are mainly inspired by the simulation study in Berg et al. (2016). The following particular case of model in (5.91) is fixed:

$$\begin{aligned} &w_{di} = \beta_0 + \beta_1 x_{di} + u_d + e_{di}; \\ &u_d \sim \mathcal{N}(0, \tau^2),\ e_{di} \sim \mathcal{N}(0, \sigma^2);\ d = 1, ..., D;\ i = 1, ..., N_D. \end{aligned} \tag{6.8}$$

The total amount of areas is $D = 10$, and for each one of the following dimensions two areas are included: $N_d = (41, 81, 161, 323, 645)$ and $n_d = (3, 5, 10, 20, 40)$, leading to a population of size $N = 2520$ and a sample of $n = 156$. At each Monte Carlo iteration, the set of covariates is generated from a $\mathcal{N}\left(\mu_x, \sigma_x^2\right)$, with $\mu_x = 3.253$ and $\sigma_x = 1.58$. The model coefficients are fixed equal to $\beta_0 = -1.62$ and $\beta_1 = 0.9$. Finally, the variance components are set according to three different scenarios: $\left(\sigma^2, \tau^2\right) = (0.6, 0.3)$ with $\phi = 0.5$, $\left(\sigma^2, \tau^2\right) =$

$(0.78, 0.12)$ with $\phi = 0.15$ and, finally, another setting having $\phi = 0.25$ and characterized by higher variances $(\sigma^2, \tau^2) = (2, 0.5)$. The frequentist properties of the estimates are evaluated with $B = 5000$ iterations and the hierarchical Bayes posterior summaries are computed on 4000 samples.

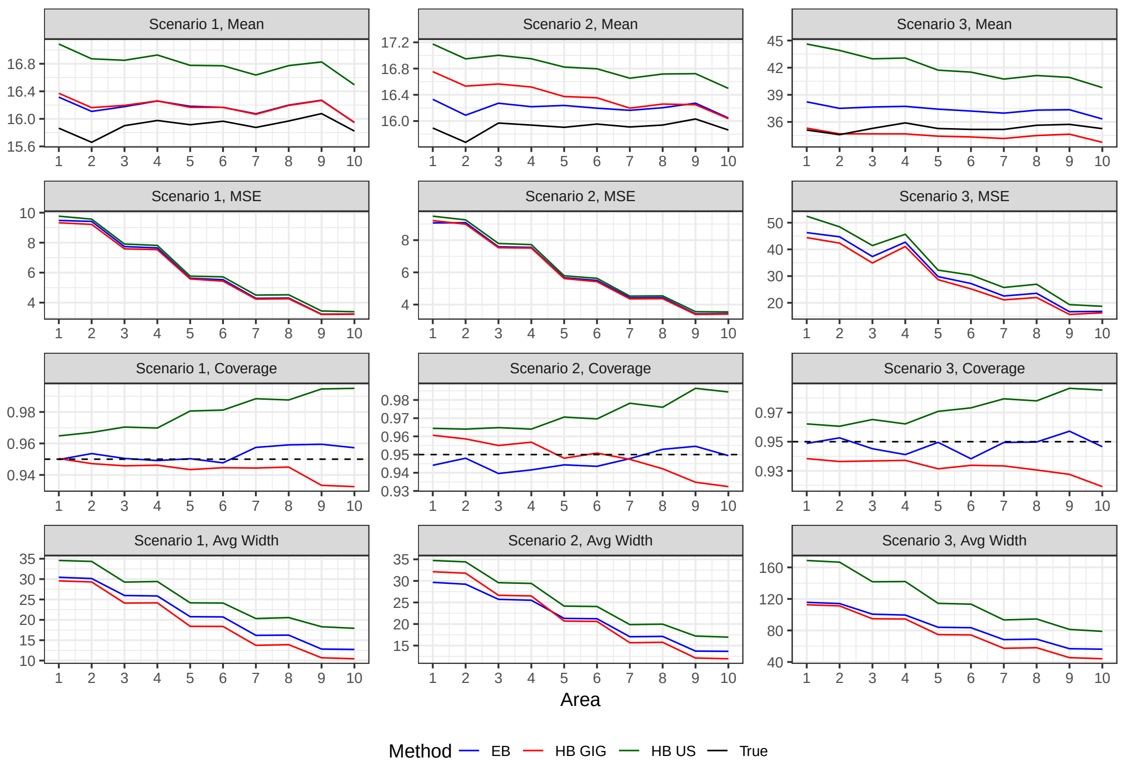

Figure 6.1: Trends of the mean estimate, RMSE, frequentist coverage and average interval width are reported for each area and for the three considered scenarios.

The area means are the target quantity to estimate and the proposals described in the previous chapter are compared. In particular, the EB methodology and the HB methodologies under GIG priors having tail parameter that assures the posterior moments finiteness and the naive uniform shrinkage prior (Molina et al., 2014) are included in the simulation study. Bias and RMSE are used to assess the performances of point estimates, whereas coverage and average width are employed to evaluate the intervals.
The main results are reported for all the areas in plots contained in figure 6.1, and summarized according to the area sample sizes in table 6.4. It emerges that the area means estimates obtained through the HB method with naive priors show a worst behaviour than the other proposals. In particular, the point estimates are characterized by a severe positive bias that does not reduce with the sample size increase and, moreover, it is enhanced by higher values of σ^2 (Scenario 3). This could be interpreted as an evidence of the moment infiniteness. A direct consequence of this bias is the increase of the RMSE of the estimator.

		Scenario 1			Scenario 2			Scenario 3		
	n_d	EB	$HB_{U.S.}$	HB_{GIG}	EB	$HB_{U.S.}$	HB_{GIG}	EB	$HB_{U.S.}$	HB_{GIG}
Bias	3	0.45	1.22	0.50	0.43	1.28	0.86	3.01	9.41	0.15
	5	0.28	0.95	0.29	0.29	1.02	0.59	2.10	7.43	-0.91
	10	0.23	0.83	0.23	0.29	0.88	0.44	2.08	6.39	-0.84
	20	0.21	0.78	0.21	0.26	0.76	0.30	1.74	5.54	-1.07
	40	0.16	0.71	0.16	0.21	0.66	0.20	1.35	4.87	-1.30
MSE	3	9.45	9.67	9.27	9.08	9.37	9.10	45.49	50.44	43.38
	5	7.69	7.86	7.56	7.57	7.76	7.52	40.00	43.55	37.98
	10	5.58	5.74	5.50	5.58	5.71	5.52	28.51	31.30	26.92
	20	4.29	4.51	4.25	4.43	4.53	4.36	23.04	26.33	21.53
	40	3.23	3.42	3.22	3.43	3.55	3.40	16.71	18.99	15.94
Coverage	3	0.95	0.97	0.95	0.95	0.96	0.96	0.95	0.96	0.94
	5	0.95	0.97	0.95	0.94	0.96	0.96	0.94	0.96	0.94
	10	0.95	0.98	0.94	0.94	0.97	0.95	0.94	0.97	0.93
	20	0.96	0.99	0.94	0.95	0.98	0.94	0.95	0.98	0.93
	40	0.96	0.99	0.93	0.95	0.98	0.93	0.95	0.99	0.92
Avg Width	3	30.30	34.45	29.43	29.45	34.57	31.96	114.89	167.70	111.88
	5	25.91	29.34	24.14	25.62	29.50	26.58	100.09	141.91	94.66
	10	20.74	24.14	18.39	21.26	24.11	20.64	83.76	113.86	74.53
	20	16.22	20.43	13.82	17.09	19.91	15.71	68.67	93.93	57.74
	40	12.78	18.11	10.56	13.69	17.08	12.03	56.53	80.08	44.79

Table 6.4: For each scenario, the estimates properties averaged with respect to areas having the same sample size n_d are reported for the different methods considered.

Moreover, the infinite posterior variance leads to an over-dispersed distribution with consequent issues in the deduced credible intervals, that are wider than the others and tend to an over-coverage, if compared to the nominal level.
On the other hand, the EB and HB estimates, with the second ones under GIG priors properly setted in order to preserve the existence of the posterior predictive distribution moments, behave similarly. The proposed HB strategy slightly outperforms the EB estimates in terms of RMSE almost in every case, and it is interesting to observe that the value of σ^2 affects the bias of the HB proposal. In fact, when σ^2 increases, the bias reduces and in the third scenario it becomes negative due to the constraints introduced to preserve the moment existence.
Comparing the interval estimators performances, the average width of the credible intervals under GIG priors are in general lower that the empirical proposal by Berg and Chandra (2014) with MSE estimated using a jackknife procedure, with the partial exception of scenario 2. On the other hand, a slight under-coverage by the Bayesian proposal emerges in

areas having higher sample sizes.

Design-based simulation: AAGIS data

To evaluate the design properties of the considered solutions for the small area framework, a design-based simulation study is carried out. The synthetic finite population of size $N = 81982$ built by sampling with replacement from the 1652 farms entering the Australian Agricultural Grazing Industries Survey (AAGIS) is largely used for this kind of simulation, especially to assess methods based on the log-transformation of data (Chandra and Chambers, 2008; Berg and Chandra, 2014; Berg et al., 2016). However, it is worth to emphasize that the population is not generated from a log-normal model.
The population is split in the 29 Australian agricultural regions and the goal of the analysis is the estimation of the area mean annual firm cost. A log-normal linear mixed model like the one in (6.8) with the logarithm of the annual firm cost as response and the logarithm of the firm size as auxiliary variable is fitted.
In each one of the $B = 1000$ iteration, a simple random sample without replacement within each strata is drawn, constituting a sample of size $n = 1686$. Then the model is estimated through the EB procedure and the HB method with GIG prior. Finally, the out of samples prediction is carried out.

	EB	HB_{GIG}
Average RRMSE	0.145	0.145
Average bias	0.011	0.011

Table 6.5: Averaged RRMSE and bias obtained with the EB and HB with GIG prior methods using the AAGIS data for design-based simulation.

The Monte Carlo relative bias and the relative RMSE of the single 29 areas are averaged in order to produce the final summary results of table 6.5. This finding basically point out the design equivalence of the EB procedure and the proposed HB strategy under GIG priors.

6.2 Real data applications

6.2.1 One-way random effect model: worker exposure data

In occupational health studies, statistical methods for the assessment of workers exposure to a particular pollutants are required. Since it is usually important to take into account both the between and the within worker variability, it is common to estimate a one way random effect ANOVA model. Moreover, the log-normality assumption for exposure concentrations is usually appropriated and the following model has been frequently considered in the literature (Lyles et al., 1997a,b):

$$w_{ij} = \log(y_{ij}) = \mu + \nu_j + \varepsilon_{ij};\; j = 1, ..., m;\; i = 1, ..., n_j; \tag{6.9}$$

where y_{ij} is the measured exposure concentration for subject j and repetition i. Therefore, the random effect ν_j is the deviation for worker j from the overall mean μ in the logarithmic scale. In this framework, the marginal expectation θ_m is interpreted as the overall mean exposure, whereas $\theta_c(\nu_j)$ is the mean worker-specific exposure.
In this example, a balanced dataset from Lyles et al. (1997a) containing styrene exposures on laminators at a boat manufacturing plant is analysed. For each one of the $m = 13$ workers, $n_j = n_g = 3$ repeated measures are executed.
In order to point out the issues related to the use of Bayesian methods with naive priors in estimating quantities in the original data scale and how these problems might be masked by the high sample size, both the whole dataset and a subset containing only the first 6 workers are analysed. The Bayesian analysis of these datasets is carried out considering the prior specifications included in the simulation study: uniform shrinkage, GIG priors and half-t.

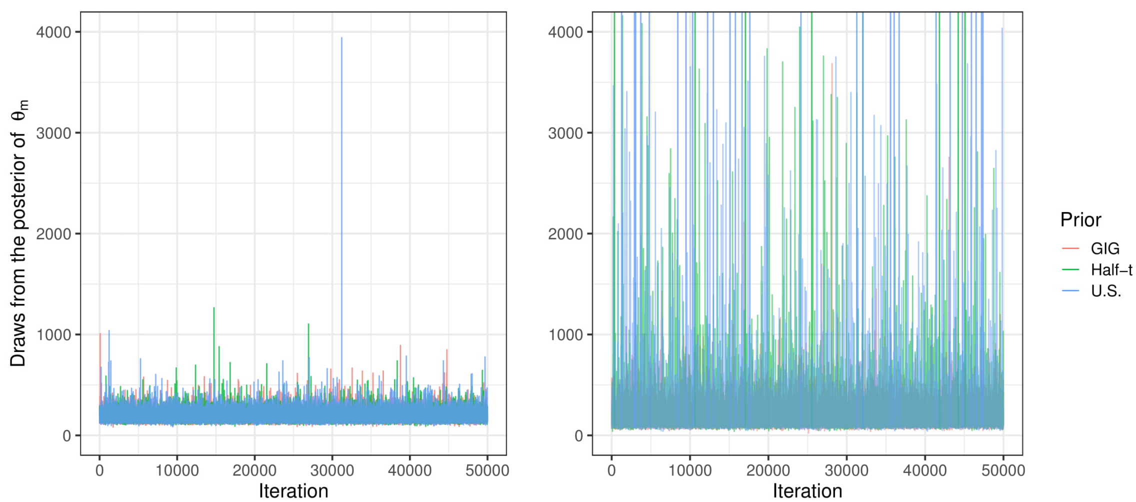

Figure 6.2: Traceplots of the overall mean θ_m under the three priors considered for the whole dataset (left) and for the sub-sample of the first 6 workers (right).

Posterior inference is carried out on 50000 iterations, after discarding the first 10000 as burn-in. The hyperparameters of the two GIG priors for the variance components are $(\lambda = 1, \delta = 0.01, \gamma = 1.92)$ for the whole dataset and $(\lambda = 1, \delta = 0.01, \gamma = 2.12)$ for the subset. Since the inferential goal is to estimate both the overall mean and the worker specific means, the most restrictive constraint between the ones on σ^2 or τ^2 is considered. It coincides with the latter and it is equal to $3 + 9/m$.
Figure 6.2 provides a first visual indication about the problems noticed using the naive priors. The traceplots of the draws from the posterior of θ_m using the whole dataset (on the left) and using a subset (on the right) are displayed: in the latter case an evident irregular behaviour is present, whereas in the first plot all the chains seem to be regular, with the exception of a moderately elevate value sampled under the uniform shrinkage prior. It must be noted that the posterior under the GIG prior does not show extreme departures in both

		Frequentist	U.S. prior		Half-t priors		GIG priors	
	Parameter	Estimate	Mean	S.D.	Mean	S.D.	Mean	S.D.
Complete dataset	μ	4.81	4.81	0.17	4.81	0.17	4.81	0.18
	σ^2	0.57	0.58	0.16	0.63	0.17	0.60	0.16
	τ^2	0.13	0.21	0.18	0.17	0.18	0.22	0.17
	θ_m	173.63	186.22	46.01	186.44	42.18	188.24	43.00
	$\theta_c(\nu_1)$	91.76	92.92	43.98	516.44	9983.02	91.56	42.63
	$\theta_c(\nu_2)$	154.52	157.91	57.63	190.13	71.82	157.40	53.27
	$\theta_c(\nu_3)$	186.61	194.70	73.32	189.42	64.61	194.73	66.23
	$\theta_c(\nu_4)$	189.43	198.32	75.06	189.84	61.44	198.00	67.46
	$\theta_c(\nu_5)$	149.04	151.77	56.03	194.00	95.31	150.18	51.76
	$\theta_c(\nu_6)$	197.03	207.34	79.23	192.96	72.91	207.73	71.92
	$\theta_c(\nu_7)$	176.56	183.18	68.80	187.21	54.73	182.30	61.45
	$\theta_c(\nu_8)$	173.79	179.77	66.47	187.45	66.69	179.24	60.61
	$\theta_c(\nu_9)$	203.20	214.39	83.92	195.24	76.89	215.14	75.96
	$\theta_c(\nu_{10})$	183.40	191.31	71.69	188.44	56.53	190.71	65.02
	$\theta_c(\nu_{11})$	233.26	252.71	106.19	218.73	135.28	252.40	94.48
	$\theta_c(\nu_{12})$	139.56	141.84	52.77	200.71	114.38	140.87	48.93
	$\theta_c(\nu_{13})$	169.38	174.73	64.04	187.45	106.48	174.28	58.48
First 6 workers	μ	4.62	4.62	0.32	4.61	0.34	4.62	0.31
	σ^2	0.27	0.34	0.16	0.37	0.19	0.35	0.16
	τ^2	0.41	0.50	0.51	0.61	0.61	0.46	0.32
	θ_m	142.19	678.87	$> 10^5$	191.60	317.70	162.34	74.94
	$\theta_c(\nu_1)$	41.08	51.59	29.73	15635.77	$> 10^5$	49.31	23.61
	$\theta_c(\nu_2)$	117.37	130.26	70.04	174.22	479.50	123.34	40.61
	$\theta_c(\nu_3)$	171.64	183.66	94.89	180.02	204.04	174.44	56.44
	$\theta_c(\nu_4)$	176.91	188.76	98.77	191.40	1109.71	179.70	57.89
	$\theta_c(\nu_5)$	109.14	121.34	60.56	213.61	4969.37	115.61	38.44
	$\theta_c(\nu_6)$	191.51	204.11	108.09	199.30	296.65	192.82	62.61

Table 6.6: Point estimates and posterior standard deviations (for Bayesian methods) of the model parameters, global expectation and conditioned expectations are reported for both the analysis carried out on the complete dataset and the reduced one.

the cases.

This graphical feature is confirmed by the posterior summaries of table 6.6. In the frequentist framework, after the REML estimation of the model parameters, the plug-in estimator for θ_m and the EB estimator for the worker-specific mean are reported, whereas posterior means and standard deviations are indicated for Bayesian methods. Observing the results for the analysis of the complete dataset, it appears that the the uniform shrinkage prior and the GIG priors lead to similar estimates for all the considered quantities, and they are higher

than the frequentist ones. This is probably due to the uniform prior on the correlation coefficient that might induce an overestimation of the lower variance component with few data. Analysing the outputs from the Bayesian model with half-t priors on the standard deviations, anomalies can be detected in the worker specific expectation estimation: both the posterior mean and standard deviation for worker appear to be unjustifiably high. All these features become more evident in the analysis carried out on the reduced dataset, in which issues for the estimation of θ_m can be noted for the uniform shrinkage prior setting too. In conclusion, it is clear how the evidences of moments existence issues vanishes when the sample size increases. Moreover, it must be stressed that, even if *reasonable* estimates could be obtained with naive priors, these values are meaningless since they attempt to numerically estimate an integral that is analytically not finite.
To conclude the illustration of the example, the `R` code used to estimate the model and standard output provided by the function `LN_hierachical` is reported. The full dataset is named `laminators` and is already included in the package `BayesLN`. The matrices `Xtilde` and `Ztilde` are created in order to obtain the posterior estimates of the overall mean and of all the group means. The target functionals of the analysis are specified through the argument `functional`. The prior parameters are fixed coherently to the instructions of section 5.3, in order to assure the existence of the posterior variance of the target functionals. An overview on the summary output of the function is reported.

```
library(BayesLN)
# Load the dataset included in the package
data("laminators")
head(laminators)

##   Worker log_Y
## 1      1 3.071
## 2      1 3.871
## 3      1 2.965
## 4      2 4.319
## 5      2 4.396
## 6      2 5.045

# Data frame for prediction
data_pred_new <- data.frame(Worker = unique(laminators$Worker))
Mod_est<-LN_hierarchical(formula_lme = log_Y~(1|Worker),
data_lme = laminators,
data_pred = data_pred_new,
functional = c("Subject","Marginal"),
order_moment = 2, nsamp = 50000, burnin = 10000)

## ----------:10.0%
## ----------:20.0%
```

```
## ----------:30.0%
## ----------:40.0%
## ----------:50.0%
## ----------:60.0%
## ----------:70.0%
## ----------:80.0%
## ----------:90.0%
## ----------:100.0%

#parameters priors
Mod_est$par_prior

##        lambda delta    gamma
## sigma2      1  0.01 1.921538
## tau2_1      1  0.01 1.921538

#posterior summaries
Mod_est$summaries

## $Iterations
## [1] "10001:50000"
##
## $Thinning
## [1] 1
##
## $Sample_size
## [1] 40000
##
## $par
##          Mean    SD Naive SE   2.5%    25%    50%    75% 97.5%     N_eff
## tau2    0.214 0.175    0.001  0.012  0.092  0.172  0.288 0.668  6159.169
## sigma2  0.597 0.158    0.001  0.356  0.484  0.575  0.684 0.969 17999.916
## beta 0  4.808 0.179    0.001  4.451  4.693  4.811  4.924 5.162 40000.000
## u 1    -0.687 0.436    0.002 -1.589 -0.979 -0.662 -0.361 0.036  5968.638
## u 2    -0.098 0.314    0.002 -0.759 -0.289 -0.082  0.098 0.513 32849.597
## u 3     0.116 0.314    0.002 -0.486 -0.083  0.097  0.308 0.776 36062.351
## u 4     0.130 0.316    0.002 -0.476 -0.070  0.110  0.324 0.796 31045.251
## u 5    -0.143 0.317    0.002 -0.813 -0.337 -0.124  0.058 0.463 29205.163
## u 6     0.172 0.319    0.002 -0.424 -0.036  0.148  0.372 0.849 26106.992
## u 7     0.050 0.310    0.002 -0.575 -0.140  0.043  0.242 0.680 38835.775
## u 8     0.032 0.312    0.002 -0.594 -0.159  0.027  0.223 0.670 39397.557
## u 9     0.205 0.325    0.002 -0.395 -0.008  0.180  0.402 0.902 22147.851
## u 10    0.093 0.313    0.002 -0.522 -0.105  0.079  0.285 0.743 34381.007
```

```
## u 11    0.360 0.348    0.002 -0.234  0.110  0.328  0.581 1.119 10951.952
## u 12   -0.216 0.323    0.002 -0.909 -0.417 -0.191  0.000 0.378 20501.317
## u 13    0.004 0.309    0.002 -0.617 -0.185  0.002  0.192 0.631 40000.000
##
## $subj
##            Mean     SD Naive SE    2.5%     25%     50%     75%   97.5%
## Subj 1   92.370 43.232    0.216  33.321  59.962  83.780 116.596 194.190
## Subj 2  157.847 53.079    0.265  77.150 121.082 150.317 185.568 282.942
## Subj 3  195.061 65.891    0.329 101.723 150.857 183.208 225.484 359.276
## Subj 4  197.963 67.458    0.337 103.013 152.157 186.203 229.658 362.055
## Subj 5  151.167 51.634    0.258  73.076 115.502 143.899 177.970 273.129
## Subj 6  206.610 70.518    0.353 108.193 158.790 193.410 239.472 384.127
## Subj 7  182.670 60.955    0.305  93.067 141.801 172.837 212.207 331.244
## Subj 8  179.489 60.205    0.301  91.663 138.595 169.711 208.770 324.718
## Subj 9  213.894 74.915    0.375 112.145 163.013 199.366 247.994 398.949
## Subj 10 190.649 64.197    0.321  98.540 147.615 179.642 221.303 346.869
## Subj 11 251.406 94.041    0.470 130.064 186.540 231.723 294.087 485.826
## Subj 12 141.012 49.170    0.246  66.489 106.436 134.279 167.427 254.726
## Subj 13 174.407 58.428    0.292  88.843 135.178 165.309 203.011 316.074
##             N_eff
## Subj 1   5317.968
## Subj 2  28214.733
## Subj 3  35533.958
## Subj 4  34815.714
## Subj 5  26462.330
## Subj 6  30214.511
## Subj 7  37812.456
## Subj 8  38952.487
## Subj 9  26464.589
## Subj 10 37902.678
## Subj 11 13340.106
## Subj 12 19092.590
## Subj 13 38697.007
##
## $marg
##              Mean     SD Naive SE    2.5%     25%     50%     75%   97.5%
## Marginal 1 188.01 42.617    0.213 127.582 159.457 181.035 207.656 290.467
##               N_eff
## Marginal 1 26567.65
```

6.2.2 Random intercept model: reading times data

Another interesting research field in which the use of log-transformed response variables in linear mixed models can be encountered is linguistics. The analysed dataset is due to Gibson and Wu (2013) and consists of a two-conditions repeated measure collection of observations of the time (in milliseconds) required to read the head noun of a Chinese clause. The following model is specified:

$$w_{ijk} = \log(y_{ijk}) = \beta_0 + \beta_1 x_i + u_j + v_k + \varepsilon_{ijk}, \tag{6.10}$$

where y_{ijk} is the reading time observed for subject $j = 1, ..., 37$, reading item $k = 1, ..., 15$ and condition $i = 1, 2$. Moreover, it is fixed $x_i = -1$ in case of subject relative, and $x_i = 1$ for object relative condition. Two random effects are included in the model in order to account for the potential correlation of observations within subject and item. The random effects are assumed independently distributed as $u_j \sim \mathcal{N}\left(0, \tau_u^2\right)$ and $v_k \sim \mathcal{N}\left(0, \tau_v^2\right)$, which are independent from the error term $\varepsilon_{ijk} \sim \mathcal{N}\left(0, \sigma^2\right)$. This model can be easily estimated using the `LN_hierarchical` function. In this case, since the matrices `Xtilde` and `Ztilde` are not specified, the covariate patterns of the sample are used for the estimation.

```
library(BayesLN)
# Load the dataset included in the package
data("ReadingTime")

#Model estimation
Mod_est_RT <- LN_hierarchical(formula_lme = log_rt ~ so  +(1|subj)+(1|item),
    data_lme = ReadingTime,
    functional = c("Marginal", "Subject"),
    nsamp = 11000, burnin = 1000)
```

In practice, the expectation conditioned on x_i and marginalized with respect both the random effect is:

$$\theta_m(x_i = \pm 1) = \exp\left\{\beta_0 \pm \beta_1 + \frac{\tau_u^2 + \tau_v^2 + \sigma^2}{2}\right\}. \tag{6.11}$$

Moreover, the expectation specific of a particular subject and item might be of interest too:

$$\theta_c(x_i, u_j, v_k) = \exp\left\{\beta_0 + x_i\beta_1 + u_j + v_k + \frac{\sigma^2}{2}\right\}, \tag{6.12}$$

as well as the expectation conditioned to only a particular random effect, e.g. integrating out only the subject:

$$\theta_c(x_i, v_k) = \exp\left\{\beta_0 + x_i\beta_1 + v_k + \frac{\tau_u^2 + \sigma^2}{2}\right\}. \tag{6.13}$$

The design matrix $\mathbf{Z}$ for the random effects is constituted by two blocks in order to define two distinct random intercepts: the elements of $\mathbf{Z}_v \in \mathbb{R}^{n\times 15}$ assume value 1 in column k

if the observation is related to the item k and 0 otherwise; on the other hand $\mathbf{Z}_u \in \mathbb{R}^{n\times 37}$ assume value 1 in column j if the observation is related to subject j and 0 otherwise.
As a consequence, the rank deficiency of $\mathbf{X}\left(\mathbf{I}-\mathbf{P}_Z\right)\mathbf{X}$ is $l=1$ and it is due to the fixed effect intercept, which is linearly dependent with respect to both $\mathbf{Z}_v$ and $\mathbf{Z}_u$.
To fix a value for the hyperparameter γ that assures the posterior moments existence for the target functionals, the different conditions related to σ^2, τ_u^2 and τ_v^2 must be computed. The easiest result to obtain is the one on σ^2, since only the maximum leverage is required. Considering the value of the condition $i)$ of proposition 5.4 deduced with the order moment $r+1=3$, in order to preserve the existence of the posterior variance, the computed optimal value $\gamma_\sigma = 1.742$ is used. Then, concerning the conditions about the random effects variances, $\mathbf{L}_v \in \mathbb{R}^{2\times 2}$ and $\mathbf{L}_u \in \mathbb{R}^{2\times 2}$ must be computed, and $\mathbf{X}_o$ coincides with $\mathbf{X}$ since the rank deficiency is due to the intercept. Given that $l=1$, the unique non-null elements of the two matrices coincide with the inverse of the first elements of $\mathbf{X}^T\left(\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^-\mathbf{C}_v(\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T\right)\mathbf{X}$ and $\mathbf{X}^T\left(\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^-\mathbf{C}_u(\mathbf{Z}^T\mathbf{Z})^-\mathbf{Z}^T\right)\mathbf{X}$, where $\mathbf{C}_v = \operatorname{diag}\left(\mathbf{I}_{15}, \mathbf{0}_{37}\right)$ and $\mathbf{C}_v = \operatorname{diag}\left(\mathbf{0}_{15}, \mathbf{I}_{37}\right)$. The deduced numerical conditions are $\gamma_{\tau,v} = 2.046$ and $\gamma_{\tau,v} = 2.434$. Therefore, the latter value is chosen since it is the most restrictive condition.
In this example, the naive Bayesian priors are not considered and the point estimates obtained under the GIG priors are compared to the REML results in table 6.7. It is possible to observe that the estimates of the basic parameters of the model are similar in both the estimation strategies, even if the estimates of the random effects variances are still slightly higher in the Bayesian proposal. This fact leads to higher estimates for the expectation marginalized with respect to the random effects $\theta_m(x_i)$. It must be highlighted that the EB estimator of the expectation conditioned with respect to the random effects cannot be used in this example, since it has been derived for the random intercept model only. Therefore, the plug-in estimators are employed for both the functionals families. Moreover, the estimation of the estimator standard error might be tricky in the frequentist context as the model complexity increases.

	REML	GIG priors	
	Estimate	Mean	S.D.
β_0	6.06	6.06	0.07
β_1	-0.04	-0.04	0.02
τ_1^2	0.18	0.22	0.06
τ_2^2	0.24	0.26	0.04
σ^2	0.52	0.52	0.02
$\theta_m(x_i = -1)$	495.97	504.19	40.53
$\theta_m(x_i = +1)$	532.78	541.60	43.27
$\theta_c(1, u_{12}, v_8)$	1396.36	1394.93	222.60
$\theta_c(1, v_8)$	769.03	760.35	79.48

Table 6.7: Point estimates and posterior standard deviations (for Bayesian method) of the model parameters and the target expectations are reported.

Chapter 7

Conclusions

In this thesis, the problem of estimating functionals in the original data scale is faced, when log-normality is assumed for data. In particular, focusing on the Bayesian framework, mathematical issues that affect the posterior distributions of such functionals are studied and new solutions are proposed for some common estimation problems which have not received any attention in the literature so far. In particular, the first part of the work concerns the quantile estimation, whereas the second part is about the estimation of conditional means under a log-normal linear mixed model.

The main difficulties in carrying out Bayesian inference on these quantities are related to the infiniteness of their posterior distribution moments. In fact, the conditions that are derived for the posterior moments existence for either the target functionals exclude the use of the most common priors for the variance in the logarithmic scale, if the usual synthesis of the posterior obtained by minimizing the quadratic loss function is desired. Therefore, it turns out that the classical inverse gamma prior for the variance or the half-t prior for the scale should be replaced by a distribution with lighter right tail. Following the proposal by Fabrizi and Trivisano (2012) in the log-normal mean estimation context, a generalized inverse Gaussian prior is assumed for the variance (or variances, in the mixed model case). Once a prior which satisfies the required existence conditions is specified, it is important to formulate a weakly informative hyperparameters setting. This task is particularly tricky because of the large amount of prior information brought by the observance of the moments existence condition. Moreover, the presence several variance components implies a more critical situation in the mixed model framework.

Concerning the prior specification issue, in the quantile estimation context, it might be interesting to develop a Bayes point estimator with optimal frequentist properties, because of the high recurrence of small samples. This is achieved by means of a numerical optimization procedure.

A further crucial step of the work is represented by the frequentist properties evaluation of the developed methods. In particular, simulation studies are illustrated to compare both the point estimators and the interval estimates to other proposals which are present in the literature. In this phase, several interesting features of the proposed methodologies are

pointed out. Concerning the point estimation problem, the Bayes estimators deduced along the work are competitive and often overcome the frequentist gold standards in terms of mean squared error. Interesting cues can be extrapolated also in the interval estimation problem. In fact, comparing the credible intervals obtained under the new prior strategy and under the classical priors which do not assure the posterior moments existence, it emerges that preserving the moments finiteness leads to more precise intervals.
Finally, the statistical methods derived and studied in the thesis are applied to real datasets taken from the literature. The results obtained under different approaches are compared and interpreted in view of the simulation study result. Moreover, the functionalities of the implemented `R` routines included in the developed package `BayesLN` are illustrated: details of the syntax and the generated outputs are reported and commented. This step could be interesting for practitioners operating in the several fields in which log-normal data are analysed.

7.1 Further developments and future work

The research reported in this thesis, jointly with the past works by Fabrizi and Trivisano (2012) and Fabrizi and Trivisano (2016), contributes to the literature with a wider understanding to Bayesian inference under log-normality assumption. In particular, thanks to the proposed formulation of the log-normal linear mixed model, a solution for a quite general class of inferential questions is provided. However, the model might be specified in a more general way than (5.42) with the addition of a structured covariance matrix in the random effect prior. If the developed model is capable to deal with a known structure for the matrix $\mathbf{D} = \oplus_{s=1}^{q}\mathbf{I}_{m_s}\tau_s^2$, the presence of additional unknown parameters $\boldsymbol{\rho}$ to control the random effects correlation might be desired. For example, the general formulation proposed by Sun et al. (2001) can be considered: $\mathbf{D}(\boldsymbol{\rho}) = \oplus_{s=1}^{q}\mathbf{B}_s(\rho_s)^{-1}\tau_s^2$, where $\mathbf{B}_s(\rho_s) = (\mathbf{A}_s - \rho_s\mathbf{C}_s)^{\zeta_s}$. In particular, $\mathbf{A}_s$ is known and positive definite and $\mathbf{C}_s$ is known and symmetric and, finally, ζ_s is a known non-negative integer. It could be interesting to verify the existence conditions of the posterior moments of the target quantities with this model formulation too.
A further appealing extension of the considered hierarchical model might be in the direction of the analysis of point-referenced data. In particular, the classical hierarchical Bayesian model for a stationary spatial process can be considered. A spatial process can be formalized as $\left\{Y(\mathbf{s}),\ \mathbf{s} \in \mathbf{D} \subset \mathbb{R}^d\right\}$, where $\mathbf{s}$ is the vector of coordinates of the spatial location and, usually, $d = 2$ or 3. It is common that the phenomenon of interest is positive and right skewed, and therefore the analysis of log-transformed data is not rare (Pilz et al., 2005; Banerjee et al., 2014).
Assuming that the sample $\mathbf{Y} = (Y(\mathbf{s}_1), ..., Y(\mathbf{s}_i), ..., Y(\mathbf{s}_n))$ is collected as realization of the spatial process, the equation of the considered model in this framework can be written as:

$$\log\left(Y(\mathbf{s}_i)\right) = W(\mathbf{s}_i) = \mu(\mathbf{s}_i) + v(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i),\ i = 1, ..., n. \tag{7.1}$$

Here $\mu(\mathbf{s}_i)$ identifies a trend that could be expressed as $\mu(\mathbf{s}_i) = \mathbf{x}(\mathbf{s_i})^T\boldsymbol{\beta}$, $v(\mathbf{s}_i)$ represents a zero-centred and stationary Gaussian spatial process and $\varepsilon(\mathbf{s}_i)$ is an uncorrelated error term. In general, the following conditional model is assumed for data:

$$\log(\mathbf{Y})\,|\boldsymbol{\beta}, \sigma^2, \tau^2, \boldsymbol{\rho} \sim \mathcal{N}_n\left(\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{I}_n + \tau^2\mathbf{K}(\boldsymbol{\rho})\right). \tag{7.2}$$

The form of the matrix $\mathbf{K}(\boldsymbol{\rho})$ is the main peculiarity of this kind of geo-statistical models. In fact, the generic entry can be expressed as $K_{ij}(\boldsymbol{\rho}) = \varphi(\mathbf{s}_i, \mathbf{s}_j, \boldsymbol{\rho})$, where $\varphi(\cdot, \cdot, \boldsymbol{\rho})$ is a suitable model for the correlation structure (e.g. exponential, spherical, Matern) that is controlled by the parameters vector $\boldsymbol{\rho}$.

Usually, the final goal of this type of analyses is to provide a prediction of the phenomenon in space. In the Bayesian framework this is performed naturally by means of the posterior predictive distribution. Given the findings of chapter 5 of this thesis, it could be of interest studying the form of such distribution to determine the moment existence condition.

Furthermore, in many applications it is frequent to deal with truncated or censored observations (Lawless, 2003; Helsel et al., 2005). In particular, this is common also for log-normally distributed data and several efforts have been addressed to develop estimation procedures which are able to incorporate these features (Balakrishnan and Mitra, 2011; Krishnamoorthy et al., 2011; Sen et al., 2019). The Bayesian framework was explored too, starting from the proper specification of the likelihood. For example, if left-censored data (e.g. measures below the instrumental detection limit in environmental sciences) the following information is available: $\mathbf{y} = (y_1, ..., y_i, ..., y_n)$ are the observed values (the detection limits are included for censored data) and $\boldsymbol{\delta} = (\delta_1, ..., \delta_i, ..., \delta_n)$ is a vector of indicators that assume value 1 if the observation is censored and 0 otherwise. If a simple log-normal distribution is assumed for data: $y_i \sim \log\mathcal{N}\left(\xi, \sigma^2\right),\ i = 1, ..., n$, then the likelihood is:

$$p(\mathbf{y}|\xi, \sigma^2) = \prod_{i=1}^{n}\left[\left(\frac{1}{y_i\sqrt{2\pi}\sigma}\exp\left\{-\frac{(\log y_i - \xi)^2}{2\sigma^2}\right\}\right)^{i-\delta_i}\left(\frac{1}{y_i}\Phi\left(\frac{\log y_i - \xi}{\sigma}\right)\right)^{\delta_i}\right]. \tag{7.3}$$

Also in this case it might be useful to investigate if the posterior moments for quantities like mean and quantiles are well defined or they require some existence condition.

In conclusion, it is interesting to develop the small area estimation model of section 5.4 and to apply it in the analysis some real dataset. This further step might be the occasion to deal with one of the major drawbacks that affects hierarchical Bayes formulation of unit-level SAE model: the feasibility of computations with high data dimensions. In particular, the theme of approximating the posterior predictive distribution when the out-of-sample observations are too many to use MCMC methods could be the target of a future research work.

Appendix A

Special Functions

A.1 Gamma Function

Even if the gamma function is a well known concept in statistics, it is important to summarize some useful relations that are employed in this work. The definition of the gamma function is based on the Eulerian integral of second kind (Andrews et al., 1999):

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} \mathrm{d}t. \tag{A.1}$$

The gamma function has poles at 0 and at all the negative integers; as a consequences the inverse gamma function $\Gamma(x)^{-1}$ has the zeroes at those points.
The gamma function has several useful relations and an important one is the *Euler reflection formula*:

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}. \tag{A.2}$$

A.2 Bessel Functions

Among the components of the wide family of the so called *Bessel functions*, in this work we are interested in the *Modified Bessel function of the second kind*, from now on called simply Bessel K function. The function of order $\nu \in \mathbb{R}$ and argument x, represents the second solution to the second order differential equation of the form:

$$x^2 \frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + x\frac{\mathrm{d}y}{\mathrm{d}x} - (x^2 + \nu^2)y = 0. \tag{A.3}$$

As it is possible to observe from Figure A.1, $K_\nu(x)$ is an exponentially decreasing function and tends to zero when the argument x is a real positive numbers with $x \to +\infty$. Moreover it approaches infinite values when $x = 0$ (Abramowitz and Stegun, 1964).
This kind of functions possess a lot of particular properties that have been used over this work:

- Integral representation, formula 3.471.9 in Gradshteyn and Ryzhik (2014):

$$\int_0^{+\infty} x^{\nu-1} \exp\left\{-\frac{\beta}{x} - \gamma x\right\} \mathrm{d}x = 2\left(\frac{\beta}{\gamma}\right)^{\frac{\nu}{2}} K_\nu\left(2\sqrt{\beta\gamma}\right). \tag{A.4}$$

- Recurrence relations:

$$K_\nu(x) = K_{-\nu}(x), \tag{A.5}$$

$$-2K'_\nu(x) = K_{\nu-1}(x) + K_{\nu+1}(x), \tag{A.6}$$

  where $K'_\nu(x)$ is the first derivative of the Bessel K function with respect to the argument.

- Asymptotic approximations:

$$K_\nu(x) \sim \frac{1}{2}\Gamma\left(|\nu|\right)\left(\frac{x}{2}\right)^{-|\nu|}, \qquad x \to 0,\ \nu \neq 0; \tag{A.7}$$

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2\nu}}\left(\frac{ex}{2\nu}\right)^{\nu}, \qquad \nu \to +\infty,\ x \neq 0; \tag{A.8}$$

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}, \qquad x \to \infty. \tag{A.9}$$

- Multiplication theorem:

$$\frac{K_\nu(\lambda z)}{\lambda^{-\nu}} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\left(\frac{(\lambda^2-1)z}{2}\right)^n K_{\nu+n}(z), \quad |\lambda^2 - 1| < 1. \tag{A.10}$$

It is possible to remark that the reported integral representation is similar to the gamma integral in (A.1) and the relationship is made explicit by the limiting form (A.7).
In the `R` software, this particular special function is implemented by the `besselK()` procedure.

A.3 Confluent Hypergeometric Function

The two independent solutions of the Kummer differential equation:

$$x\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + (b-x)\frac{\mathrm{d}y}{\mathrm{d}x} - ay = 0, \tag{A.11}$$

are named *Confluent Hypergeometric Functions*. It is frequent to meet this family in dealing with special function since they include as particular case a lot of other functions.
One of the main issues in handling the confluent hypergeometric functions is the confusing notation that is present in the literature. In fact, it is possible to distinguish:

- The Kummer's M confluent hypergeometric function: $M(a,b;x) = \Phi(a,b;x) = {}_1F_1(a,b;x)$ and is implemented in the `R` software by the function `kummerM()`.

- The Kummer's (but also Tricomi's) U confluent hypergeometric function: $U(a,b;x) = \Psi(a,b;x)$; that could be written also as a sum of two Kummer's M.

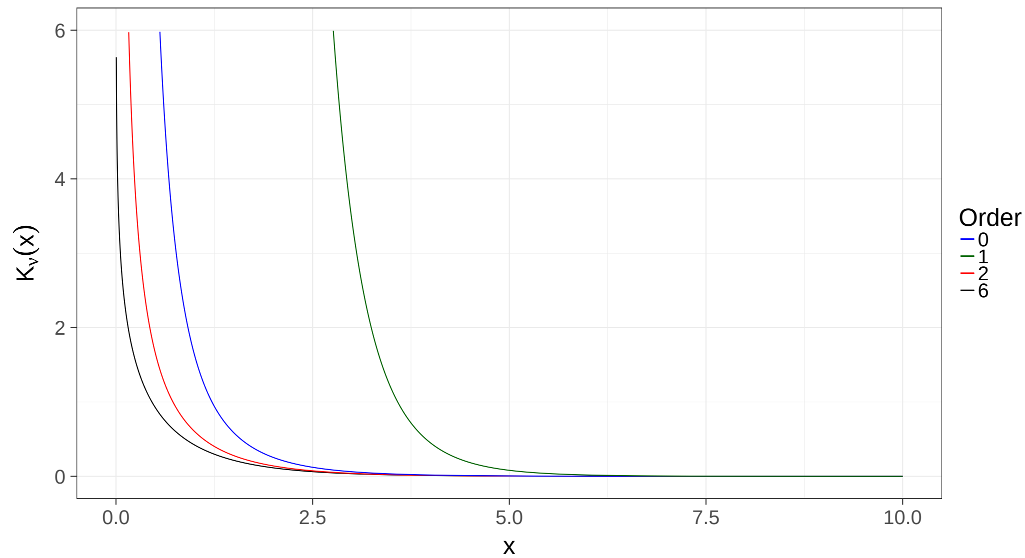

Figure A.1: Curve of the Bessel K function with different orders ν.

A.4 Parabolic Cylinder Function

The solutions of the differential equation:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} + \left(ax^2 + bx + c\right) y = 0, \tag{A.12}$$

are the *Parabolic Cylinder Functions* (Abramowitz and Stegun, 1964). In the literature it is possible to find two kinds of notation. The link between these conventions is the following: $D_\nu(x) = U\left(-\frac{1}{2} - \nu, x\right)$.

The parabolic cylinder functions have integral representations that are quite common in probability:

$$\int_0^{+\infty} x^{\nu-1} e^{-\beta x^2 - \gamma x} \mathrm{d}x = \frac{\Gamma(\nu)}{(2\beta)^{\nu/2}} \exp\left\{\frac{\gamma^2}{8\beta}\right\} D_{-\nu}\left(\frac{\gamma}{\sqrt{2\beta}}\right), \quad \begin{matrix} \beta > 0 \\ \nu > 0 \end{matrix}, \tag{A.13}$$

$$\int_{-\infty}^{+\infty} (ix)^\nu e^{-\beta^2 x^2 - iqx} \mathrm{d}x = \frac{\sqrt{\pi}}{2^{-\frac{\nu}{2}} \beta^{-\nu-1}} \exp\left\{-\frac{q^2}{8\beta^2}\right\} D_\nu\left(\frac{q}{\beta\sqrt{2}}\right), \quad \begin{matrix} \beta > 0 \\ \nu > -1 \end{matrix}; \tag{A.14}$$

and they are, respectively, eqn. 3.462.1 and 3.462.3 in Gradshteyn and Ryzhik (2014).
Besides, the parabolic cylinder function is strictly connected with the confluent hypergeometric function, and, in particular, it is often use the following relationship with the Kummer's M:

$$D_\nu(x) = 2^{\frac{\nu}{2}} e^{-\frac{x^2}{4}} \left[\frac{\sqrt{\pi}}{\Gamma\left(\frac{1-\nu}{2}\right)} M\left(-\frac{\nu}{2}, \frac{1}{2}; \frac{x^2}{2}\right) - \frac{\sqrt{2\pi}x}{\Gamma\left(-\frac{\nu}{2}\right)} M\left(\frac{1-\nu}{2}, \frac{3}{2}; \frac{x^2}{2}\right)\right], \tag{A.15}$$

from equation 19.12.3 in Abramowitz and Stegun (1964).

Appendix B

Useful Distributions

B.1 Generalized Hyperbolyc Distribution

In general, a *Gaussian variance-mean mixture* is obtained considering the random variable X that has a Gaussian conditional distribution:

$$X|W = w \sim \mathcal{N}\left(\mu + \beta w, w\right), \tag{B.1}$$

where μ and β are constants and W is a random variable with positive support that follows a given probability distribution, called also *mixing distribution*. This particular setting produces distributions that have common interesting properties that are collected, for example, in the book by Paolella (2007).
The generalized hyperbolic (GH) distribution was introduced in a paper by Barndorff-Nielsen (1977) and it is obtained considering a normal variance-mean mixture with a generalized inverse Gaussian distribution as mixing distribution (see section 2.1). So:

$$\left.\begin{array}{l} X|W = w \sim \mathcal{N}\left(\mu + \beta w, w\right) \\ W \sim GIG\left(\lambda, \delta, \gamma\right) \end{array}\right\} \implies X \sim GH(\lambda, \alpha, \beta, \delta, \gamma),$$

where $\alpha^2 = \gamma^2 + \beta^2$ and X is a real valued random variable with the following probability density function:

$$f(x) = \frac{\left(\frac{\gamma}{\delta}\right)}{\sqrt{2\pi} K_\lambda(\delta\gamma)} \frac{K_{\lambda-\frac{1}{2}}\left(\gamma\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{\frac{1}{2}-\lambda}} \exp\left\{\beta(x-\mu)\right\}, \tag{B.2}$$

with the following restrictions on the parameters:

- $\lambda = 0 \implies |\beta| < \alpha,\ \delta > 0;$
- $\lambda > 0 \implies |\beta| < \alpha,\ \delta \geq 0;$
- $\lambda < 0 \implies |\beta| \leq \alpha,\ \delta > 0.$

To have a detailed discussion about the mathematical properties of the GH distribution it is worth to look up in Bibby and Sørensen (2003) and in Hammerstein (2010). The expectation of the GH distribution is:

$$\mathbb{E}\left[X\right] = \mu + \frac{\beta\delta K_{\lambda+1}(\gamma\delta)}{\gamma K_{\lambda}(\gamma\delta)}; \tag{B.3}$$

and the moment generating function:

$$\mathbb{M}[r] = \exp\left\{\mu t\right\}\left(\frac{\gamma^2}{\alpha^2-(\beta+r)^2}\right)^{\frac{\lambda}{2}}\frac{K_{\lambda}\left(\delta\sqrt{\alpha^2-(\beta+r)^2}\right)}{K_{\lambda}(\delta\gamma)},\ \ |\beta+r|<\alpha. \tag{B.4}$$

Examples of well known distributions that are particular cases of the GH distribution are the *hyperbolic asymmetric t* distribution, the *asymmetric Laplace* distribution, the *Student's t* distribution.

B.1.1 The Multivariate GH distribution

It is possible to generalize idea of normal variance-mean mixture distribution in the multivariate field. In particular, a random vector $\mathbf{X} \in \mathbb{R}^d$ is said to have a normal mean variance mixture distribution if:

$$\mathbf{X} = \boldsymbol{\mu} + W\boldsymbol{\beta} + \sqrt{W}A\mathbf{Z}, \tag{B.5}$$

where $\boldsymbol{\beta}, \boldsymbol{\mu} \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ such that $\Delta = AA^T$ is positive definite, $\mathbf{Z} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{I}_d\right)$ and W is a real valued positive random variable independent with respect to $\mathbf{Z}$. The multivariate GH of dimension d (MVGH$_d$) was introduced in the paper by Barndorff-Nielsen (1977) and it is obtained in the following way:

$$\left.\begin{array}{l} W \sim GIG\left(\lambda, \delta, \sqrt{\alpha^2 - \boldsymbol{\beta}^T\Delta\boldsymbol{\beta}}\right) \\ \mathbf{X}|W = w \sim MVN_d\left(\boldsymbol{\mu} + w\boldsymbol{\beta}\Delta, w\Delta\right) \end{array}\right\} \implies \mathbf{X} \sim MVGH_d\left(\lambda, \alpha, \boldsymbol{\mu}, \Delta, \delta, \boldsymbol{\beta}\right),$$

where $\lambda \in \mathbb{R}$, $\alpha > 0$, $\delta \geq 0$ and $0 < \sqrt{\boldsymbol{\beta}^T\Delta\boldsymbol{\beta}} < \alpha$. The resultant density function is:

$$\begin{aligned} f(\mathbf{x}) =& \frac{\left(\alpha^2 - \boldsymbol{\beta}^T\Delta\boldsymbol{\beta}\right)^{\frac{\lambda}{2}}}{(2\pi)^{\frac{d}{2}}\sqrt{|\Delta|}\alpha^{\lambda-\frac{1}{2}}}\times \\ &\times\frac{K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2+(\mathbf{x}-\boldsymbol{\beta})^T\Delta^{-1}(\mathbf{x}-\boldsymbol{\beta})}\right)e^{-\boldsymbol{\beta}^T(\mathrm{x}-\boldsymbol{\beta})}}{K_{\lambda}\left(\delta\sqrt{\alpha^2-\boldsymbol{\beta}^T\Delta\boldsymbol{\beta}}\right)\left(\sqrt{\delta^2+(\mathbf{x}-\boldsymbol{\beta})^T\Delta^{-1}(\mathbf{x}-\boldsymbol{\beta})}\right)^{\frac{d}{2}-\lambda}}. \end{aligned} \tag{B.6}$$

Appendix C

Bayesian point estimation

In the framework of decision theory, an important introduced tool is an evaluation criterion of the decision, usually named *loss*. Basically, point estimation can be considered as a particular decision problem in which the so called decision space $d \in \mathcal{D}$ might coincide with the parameter space Θ, with the true value of the parameter $\theta \in \Theta$. The true value of the decision is a function of θ, say $h(\theta)$ and it is proposed an estimate $\delta(\mathbf{x})$, based on the observations $\mathbf{x} \in \mathbf{X}$. Now it is possible to define:

Definition C.1. *A loss function L is a non-negative function:*

$$L(\theta, \delta(\mathbf{x})) \geq 0, \quad \forall\theta,\ \delta(\cdot),$$

which takes value 0 if the estimate coincides with the true value (no loss in that case):

$$L(\theta, \delta(\mathbf{x})) = 0, \quad \forall\theta.$$

A uniform minimization of $L(\theta, d)$ is impossible in most cases, therefore it is usually adopted the $frequentist$ criterion that minimizes the average loss, called also $risk$:

$$\begin{aligned} R(\theta, \delta) &= \mathbb{E}_\theta\big[L(\theta, \delta(\mathbf{x}))\big] \\ &= \int_{\mathbf{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta) d\mathbf{x}, \end{aligned} \tag{C.1}$$

where $f(\mathbf{x}|\theta)$ is the distribution of the observation given the parameters.
From the bayesian point of view (Robert, 2007), since data $\mathbf{x}$ is known and the parameter θ is not, the rule is to integrate with respect to the parameter space Θ. In this way is obtained the so called *posterior expected loss*:

$$\begin{aligned} \rho(\pi, d|\mathbf{x}) &= \mathbb{E}^\pi\big[L(\theta, d)|\mathbf{x}\big] \\ &= \int_{\Theta} L(\theta, d)\pi(\theta|\mathbf{x}) d\theta, \end{aligned} \tag{C.2}$$

where $\pi(\theta|x)$ is the posterior distribution of θ conditioned with respect to the observations x and it is the quantity that averages the error. Considering a prior distribution $\pi(\theta)$, it is possible to define the *integrated risk*, i.e. the frequentist risk averaged over the Θ space:

$$\begin{aligned} r(\pi,\delta) &= \mathbb{E}^{\pi}\big[R(\theta,\delta)\big] \\ &= \int_{\Theta}\int_{\mathbf{X}} L(\theta,d(\mathbf{x}))f(\mathbf{x}|\theta)d\mathbf{x}\pi(\theta)d\theta. \end{aligned} \tag{C.3}$$

In this case a number is associated to each estimator, therefore it is possible to order a set of estimators by a direct comparison. An important result that holds is the following:

Theorem C.1. *An estimator that minimizes the integrated risk $r(\pi,\delta)$ can be obtained by selecting for each $\mathbf{x}$ the value $\delta(\mathbf{x})$ which minimizes the posterior expected loss $\rho(\pi,d|\mathbf{x})$, since:*

$$r(\pi,d) = \int_{\mathbf{X}} \rho(\pi,d(\mathbf{x})|\mathbf{x})m(\mathbf{x})dx,$$

where $m(\mathbf{x})$ is the function such that:

$$f(\mathbf{x}|\theta)\pi(\theta) = \pi(\theta|\mathbf{x})m(\mathbf{x}).$$

Now it is possible to provide a definition of Bayes estimator:

Definition C.2. *A Bayes estimator associated with a prior distribution π and a loss function L is any estimator δ^{π} which minimizes $r(\pi,\delta)$. For each x it is given by:*

$$\delta^{\pi}(\mathbf{x}) = \min_{d} \rho(\pi,d|\mathbf{x}).$$

Besides, the value $r(\pi) = r(\pi,\delta^{\pi})$ is called Bayes risk.

If no information is available to specify a proper loss function for the problem, a *classical loss function* could be employed. The most common one is the quadratic loss, defined as:

$$L(\theta,d) = (\theta-d)^2. \tag{C.4}$$

Even if it extremely penalizes large deviations, it has some advantages and it can be considered as a Taylor expansion approximation to other symmetric losses. Besides, the Bayes estimators under quadratic loss is the posterior mean, even if it is not the only loss function with this property. In effect, since the posterior expected quadratic loss is:

$$\mathbb{E}^{\pi}\big[(\theta-d)^2|\mathbf{x}\big] = \mathbb{E}^{\pi}\big[\theta^2|\mathbf{x}\big] - 2d\mathbb{E}^{\pi}\big[\theta|\mathbf{x}\big] + d^2,$$

then it has the minimum value at:

$$\delta^{\pi}(\mathbf{x}) = \mathbb{E}^{\pi}\big[\theta|\mathbf{x}\big]. \tag{C.5}$$

Another useful loss function is the relative quadratic loss:

$$L(\theta, d) = \left(\frac{\theta - d}{\theta}\right)^2 = \left(1 - \frac{d}{\theta}\right)^2, \tag{C.6}$$

with:

$$\mathbb{E}^{\pi}\left[\left(1 - \frac{d}{\theta}\right)^2 \middle| \mathbf{x}\right] = 1 - 2d\mathbb{E}^{\pi}\left[\frac{1}{\theta}\middle|\mathbf{x}\right] + d^2\mathbb{E}^{\pi}\left[\frac{1}{\theta^2}\middle|\mathbf{x}\right].$$

This quantity is minimized by:

$$\delta^{\pi}(x) = \frac{\mathbb{E}^{\pi}\left[\theta^{-1}|\mathbf{x}\right]}{\mathbb{E}^{\pi}\left[\theta^{-2}|\mathbf{x}\right]}. \tag{C.7}$$

Appendix D

Additional figures and tables

Table D.1: Root MSE and relative bias of estimators for θ_p with respect to different sample sizes n and quantiles p, with $\sigma^2 = 0.5$. The Bayes estimator $\hat{\theta}_p^B$ is the estimator under relative quadratic loss for the median and the one under quadratic loss for the others.

		Root MSE								
		p								
n	Method	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
	$\hat{\theta}_p^*$	0.015	0.040	0.067	0.158	0.410	1.064	2.511	4.197	11.002
	$\hat{\theta}_p^*$	0.017	0.027	0.035	0.053	0.086	0.138	0.212	0.274	0.443
11	$\hat{Q}_p^7$	0.089	0.071	0.074	0.079	0.113	0.178	0.324	0.458	0.984
11	$\tilde{\theta}_p$	0.039	0.048	0.053	0.064	0.090	0.161	0.310	0.468	1.013
	$\hat{Q}_p$	0.036	0.045	0.051	0.063	0.086	0.147	0.279	0.423	0.935
	$\hat{\theta}_p^{Ba}$	0.034	0.043	0.048	0.061	-	0.155	0.285	0.419	0.875
	$\hat{\theta}_p^{Bn}$	0.035	0.045	0.051	0.063	0.087	0.147	0.271	0.395	0.785
	$\hat{\theta}_p^*$	0.012	0.019	0.025	0.038	0.062	0.100	0.153	0.198	0.321
	$\hat{Q}_p^7$	0.057	0.049	0.049	0.056	0.080	0.136	0.247	0.371	0.813
21	$\tilde{\theta}_p$	0.026	0.032	0.036	0.044	0.064	0.114	0.216	0.322	0.676
	$\hat{Q}_p$	0.025	0.031	0.036	0.044	0.062	0.109	0.206	0.309	0.658
	$\hat{\theta}_p^{Ba}$	0.024	0.031	0.035	0.043	-	0.112	0.208	0.306	0.632
	$\hat{\theta}_p^{Bn}$	0.025	0.032	0.036	0.043	0.062	0.108	0.198	0.289	0.576
	$\hat{\theta}_p^*$	0.008	0.012	0.016	0.025	0.040	0.064	0.099	0.128	0.207
	$\hat{Q}_p^7$	0.032	0.028	0.029	0.034	0.051	0.086	0.165	0.247	0.596
51	$\tilde{\theta}_p$	0.016	0.020	0.022	0.028	0.040	0.072	0.135	0.200	0.411
	$\hat{Q}_p$	0.015	0.020	0.022	0.028	0.040	0.071	0.133	0.197	0.408
	$\hat{\theta}_p^{Ba}$	0.015	0.019	0.022	0.028	-	0.072	0.134	0.196	0.401
	$\hat{\theta}_p^{Bn}$	0.015	0.020	0.023	0.028	0.040	0.070	0.131	0.191	0.383
		Relative Bias								
	$\hat{\theta}_p^*$	-0.004	-0.006	-0.007	-0.011	-0.018	-0.029	-0.045	-0.058	-0.094
	$\hat{Q}_p^7$	0.070	0.045	0.037	0.026	0.014	-0.007	-0.097	-0.163	-0.783
11	$\tilde{\theta}_p$	0.010	0.011	0.011	0.010	0.009	0.011	0.021	0.037	0.110
	$\hat{Q}_p$	0.000	-0.004	-0.007	-0.012	-0.019	-0.024	-0.019	-0.005	0.068
	$\hat{\theta}_p^{Ba}$	0.003	0.002	0.002	0.005	-	-0.009	-0.047	-0.077	-0.138
	$\hat{\theta}_p^{Bn}$	-0.007	-0.010	-0.011	-0.013	-0.023	-0.043	-0.088	-0.130	-0.258
	$\hat{\theta}_p^*$	-0.002	-0.003	-0.004	-0.006	-0.010	-0.016	-0.024	-0.031	-0.050
	$\hat{Q}_p^7$	0.043	0.026	0.020	0.013	0.007	-0.006	-0.054	-0.136	-0.543
21	$\tilde{\theta}_p$	0.005	0.005	0.005	0.005	0.005	0.006	0.011	0.019	0.055
	$\hat{Q}_p$	-0.001	-0.003	-0.004	-0.007	-0.010	-0.012	-0.007	0.002	0.046
	$\hat{\theta}_p^{Ba}$	0.001	0.000	0.000	0.002	-	-0.004	-0.024	-0.038	-0.067
	$\hat{\theta}_p^{Bn}$	-0.006	-0.008	-0.008	-0.006	-0.011	-0.024	-0.053	-0.082	-0.174
	$\hat{\theta}_p^*$	-0.001	-0.001	-0.002	-0.002	-0.004	-0.006	-0.010	-0.013	-0.020
	$\hat{Q}_p^7$	0.021	0.011	0.008	0.006	0.003	-0.002	-0.023	-0.054	-0.265
51	$\tilde{\theta}_p$	0.002	0.002	0.002	0.002	0.002	0.002	0.004	0.007	0.019
	$\hat{Q}_p$	0.000	-0.001	-0.002	-0.003	-0.004	-0.004	-0.002	0.001	0.020
	$\hat{\theta}_p^{Ba}$	0.000	0.000	0.000	0.001	-	-0.002	-0.010	-0.016	-0.029
	$\hat{\theta}_p^{Bn}$	-0.002	-0.003	-0.003	-0.002	-0.004	-0.012	-0.030	-0.048	-0.108

Table D.2: Root MSE and relative bias of estimators for θ_p with respect to different sample sizes n and quantiles p, with $\sigma^2 = 2$. The Bayes estimator $\hat{\theta}_p^B$ is the estimator under relative quadratic loss for the median and the one under quadratic loss for the others.

		Root MSE								
		p								
n	Method	0.01	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.99
11	$\hat{\theta}_p^*$	0.002	0.007	0.011	0.026	0.067	0.174	0.410	0.685	1.796
	$\hat{Q}_p^7$	0.027	0.029	0.034	0.051	0.108	0.276	0.733	1.692	3.556
	$\tilde{\theta}_p$	0.009	0.016	0.021	0.037	0.082	0.240	0.765	1.599	6.683
	$\hat{Q}_p$	0.008	0.014	0.018	0.032	0.069	0.178	0.501	1.004	2.856
	$\hat{\theta}_p^{Ba}$	0.009	0.015	0.021	0.039	-	0.228	0.644	1.276	5.194
	$\hat{\theta}_p^{Bn}$	0.007	0.012	0.017	0.032	0.069	0.178	0.476	0.874	2.802
21	$\hat{\theta}_p^*$	0.002	0.005	0.008	0.019	0.049	0.128	0.301	0.503	1.320
	$\hat{Q}_p^7$	0.015	0.017	0.020	0.032	0.071	0.190	0.531	1.008	3.534
	$\tilde{\theta}_p$	0.005	0.009	0.013	0.024	0.055	0.158	0.479	0.955	3.535
	$\hat{Q}_p$	0.005	0.008	0.012	0.022	0.050	0.135	0.393	0.778	2.852
	$\hat{\theta}_p^{Ba}$	0.005	0.009	0.013	0.024	-	0.157	0.454	0.892	3.359
	$\hat{\theta}_p^{Bn}$	0.004	0.008	0.012	0.022	0.050	0.134	0.371	0.692	2.242
51	$\hat{\theta}_p^*$	0.001	0.003	0.005	0.012	0.032	0.083	0.196	0.328	0.860
	$\hat{Q}_p^7$	0.007	0.008	0.011	0.018	0.043	0.116	0.339	0.668	3.007
	$\tilde{\theta}_p$	0.003	0.005	0.008	0.015	0.034	0.097	0.282	0.546	1.877
	$\hat{Q}_p$	0.003	0.005	0.007	0.014	0.032	0.090	0.262	0.507	1.760
	$\hat{\theta}_p^{Ba}$	0.003	0.005	0.008	0.015	-	0.097	0.279	0.537	1.862
	$\hat{\theta}_p^{Bn}$	0.003	0.007	0.009	0.016	0.032	0.090	0.253	0.475	1.523
		Relative Bias								
11	$\hat{\theta}_p^*$	-0.001	-0.003	-0.004	-0.010	-0.027	-0.071	-0.167	-0.279	-0.731
	$\hat{Q}_p^7$	0.019	0.018	0.018	0.020	0.024	0.030	-0.089	-0.089	-2.263
	$\tilde{\theta}_p$	0.003	0.005	0.006	0.009	0.015	0.040	0.139	0.315	1.489
	$\hat{Q}_p$	-0.001	-0.002	-0.003	-0.009	-0.028	-0.068	-0.121	-0.137	-1.640
	$\hat{\theta}_p^{Ba}$	0.004	0.006	0.008	0.013	-	0.022	0.009	0.026	0.502
	$\hat{\theta}_p^{Bn}$	0.000	-0.002	-0.003	-0.009	-0.028	-0.085	-0.245	-0.457	-1.419
21	$\hat{\theta}_p^*$	-0.001	-0.001	-0.002	-0.006	-0.015	-0.039	-0.092	-0.154	-0.403
	$\hat{Q}_p^7$	0.010	0.009	0.009	0.010	0.012	0.010	-0.047	-0.214	-1.518
	$\tilde{\theta}_p$	0.001	0.002	0.003	0.004	0.008	0.020	0.068	0.152	0.685
	$\hat{Q}_p$	0.000	-0.001	-0.002	-0.006	-0.015	-0.034	-0.052	-0.043	0.171
	$\hat{\theta}_p^{Ba}$	0.002	0.002	0.003	0.006	-	0.015	0.022	0.053	0.416
	$\hat{\theta}_p^{Bn}$	-0.001	-0.002	-0.004	-0.006	-0.016	-0.049	-0.153	-0.298	-0.992
51	$\hat{\theta}_p^*$	0.000	-0.001	-0.001	-0.002	-0.006	-0.016	-0.038	-0.064	-0.168
	$\hat{Q}_p^7$	0.005	0.004	0.004	0.004	0.005	0.005	-0.019	-0.067	-0.571
	$\tilde{\theta}_p$	0.001	0.001	0.001	0.002	0.003	0.008	0.026	0.057	0.250
	$\hat{Q}_p$	0.000	0.000	-0.001	-0.002	-0.006	-0.013	-0.018	-0.011	0.099
	$\hat{\theta}_p^{Ba}$	0.001	0.001	0.001	0.002	-	0.007	0.013	0.029	0.183
	$\hat{\theta}_p^{Bn}$	-0.001	-0.002	-0.004	-0.006	-0.006	-0.024	-0.083	-0.170	-0.611

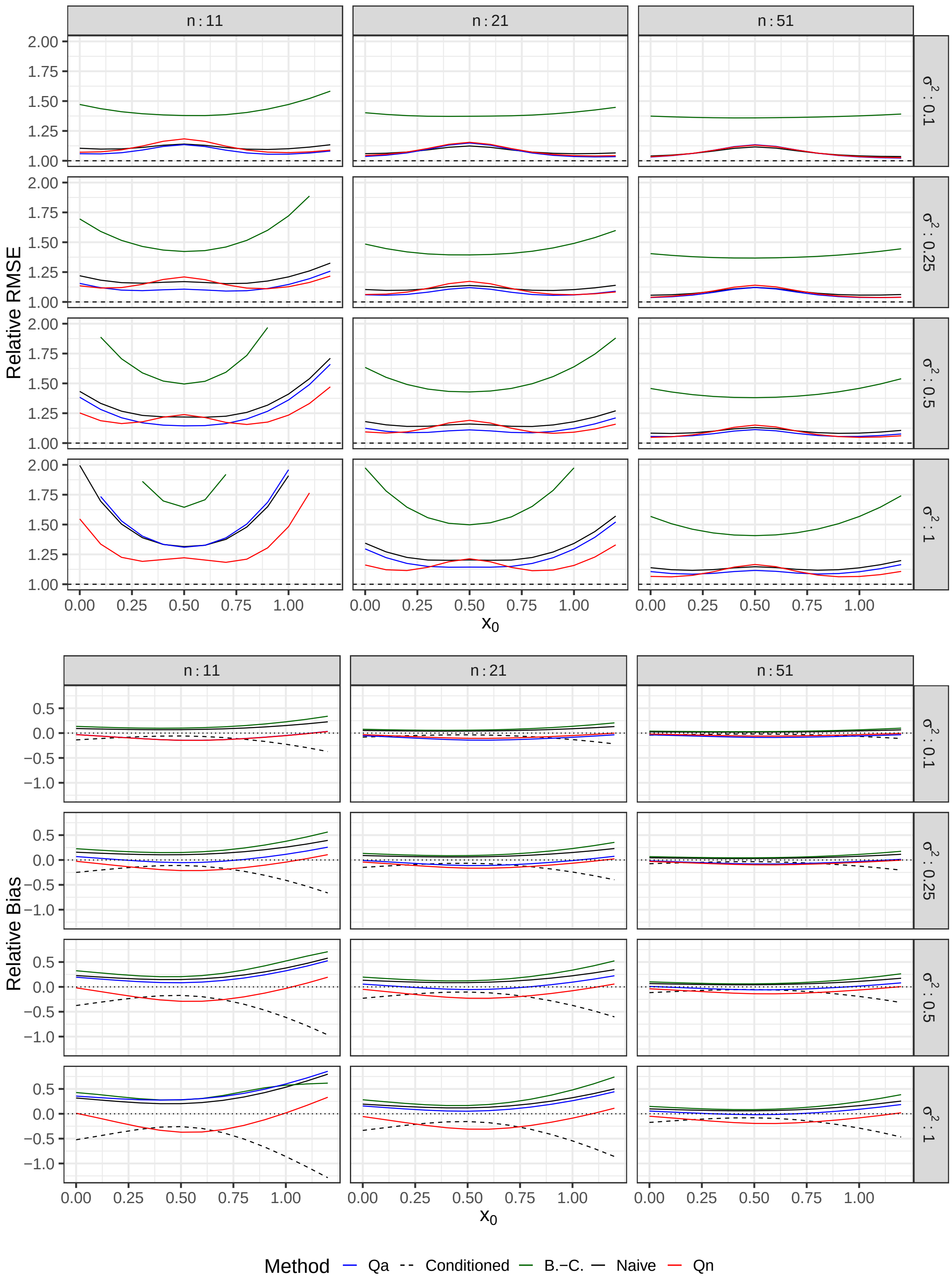

Figure D.1: Relative RMSE and relative bias of various estimators of the target quantity $\theta_p(\mathbf{x}_0)$, with $p = 0.25$.

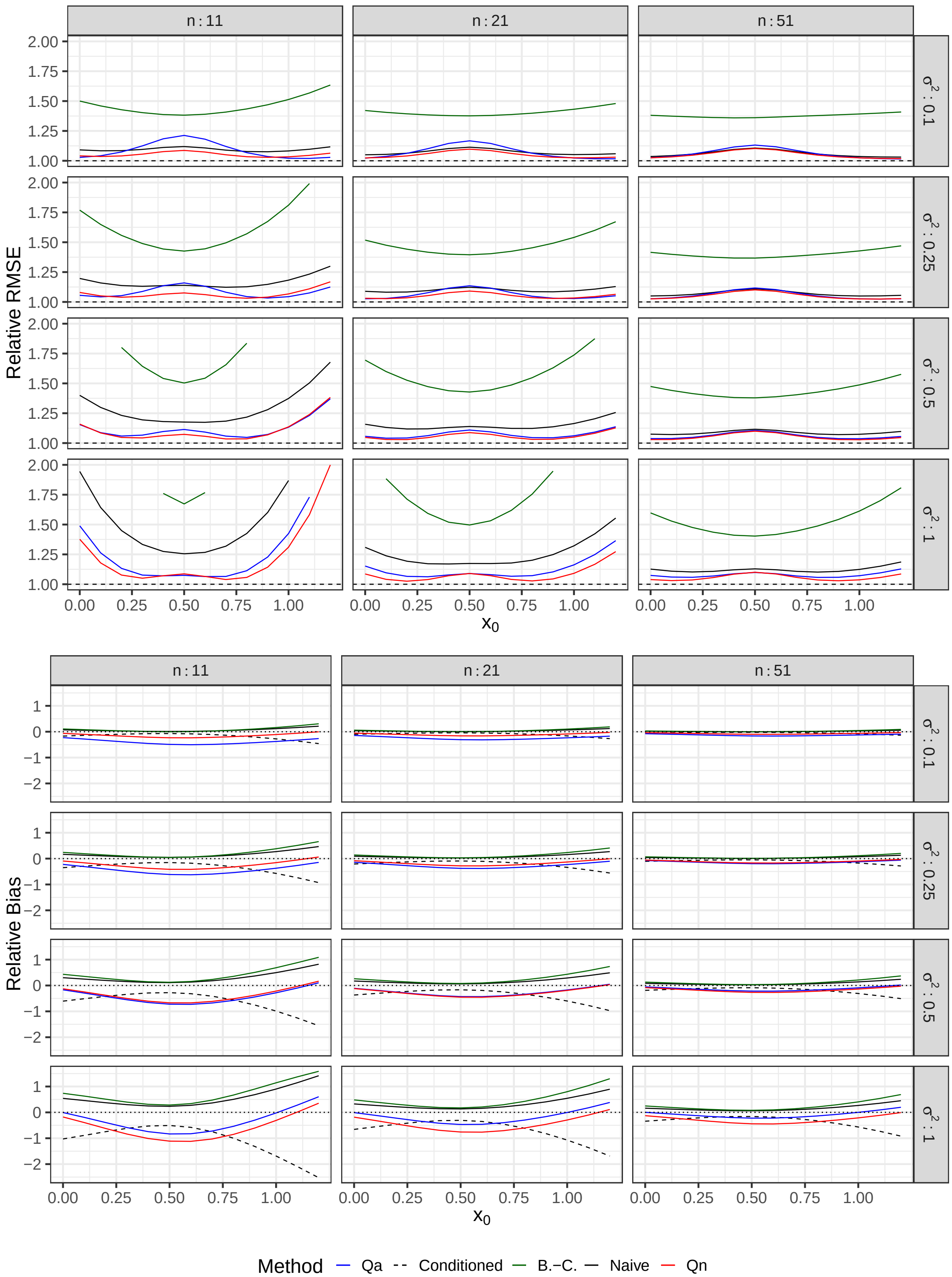

Figure D.2: Relative RMSE and relative bias of various estimators of the target quantity $\theta_p(\mathbf{x}_0)$, with $p = 0.75$.

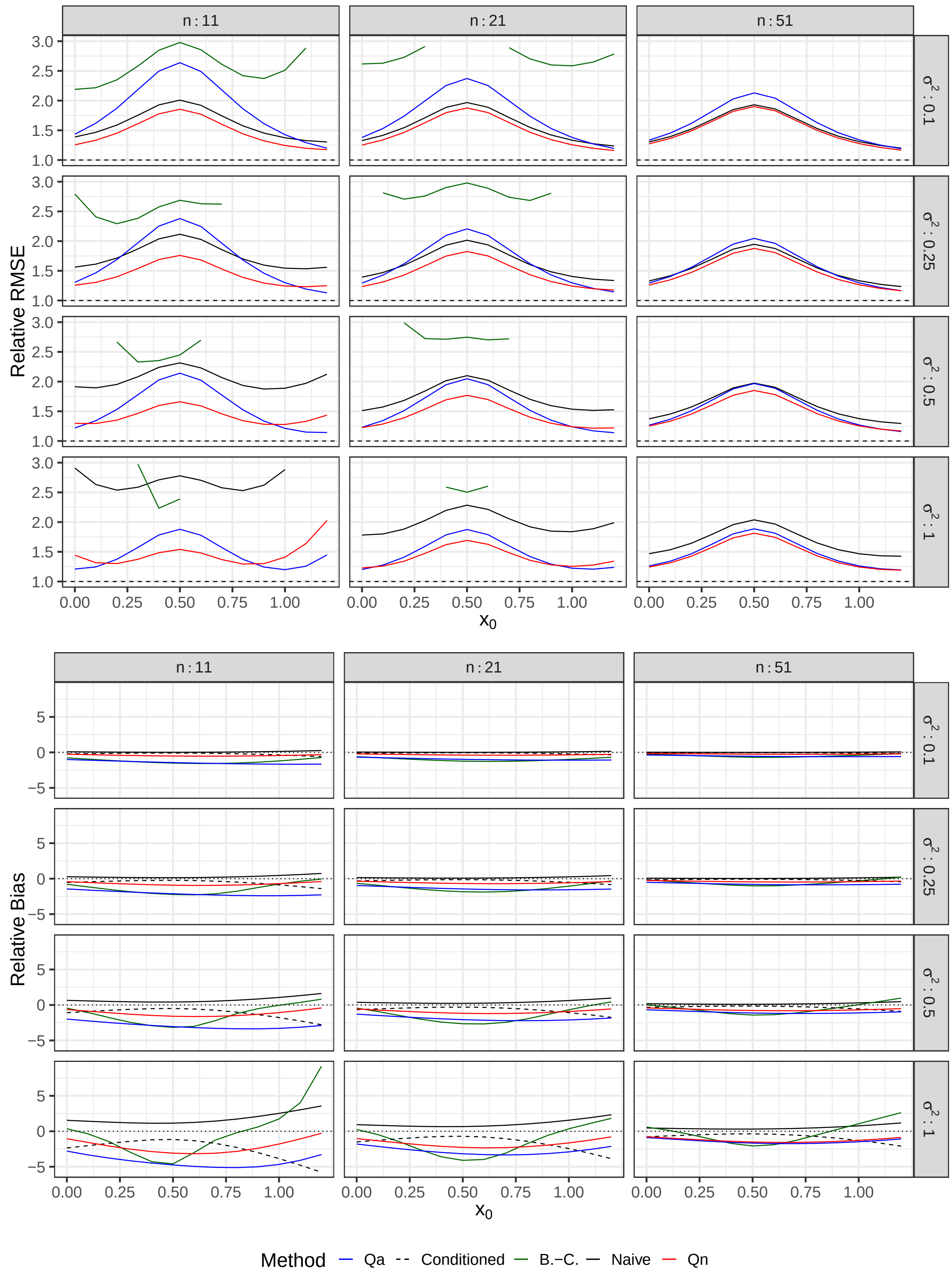

Figure D.3: Relative RMSE and relative bias of various estimators of the target quantity $\theta_p(\mathbf{x}_0)$, with $p = 0.99$.

						Global predictor				Conditioned Predictor			
n_g	m	ϕ	σ^2	Scen.	θ_m	Bias	RMSE	Cov.	Wid.	RABias	RRMSE	Aco	AWi
2	10	0.5	0.5	1	1.45	0.14	0.44	0.95	1.82	0.18	0.53	0.93	2.33
			1	2	2.12	0.17	0.92	0.94	3.67	0.32	0.91	0.91	4.60
			2	3	4.48	-0.36	2.52	0.86	9.14	0.53	1.70	0.87	11.59
		1	0.5	4	1.65	0.16	0.62	0.93	2.45	0.27	0.69	0.92	2.96
			1	5	2.72	0.09	1.42	0.90	5.29	0.52	1.35	0.90	6.45
			2	6	7.39	-1.63	4.79	0.75	14.74	1.03	3.28	0.87	19.56
		2	0.5	7	2.12	0.15	1.03	0.91	3.85	0.36	0.89	0.92	4.17
			1	8	4.48	-0.44	2.81	0.82	9.34	0.77	2.02	0.90	11.18
			2	9	20.09	-9.79	14.99	0.52	31.81	1.90	7.68	0.87	50.77
	20	0.5	0.5	10	1.45	0.08	0.29	0.96	1.19	0.17	0.51	0.92	2.14
			1	11	2.12	0.12	0.64	0.95	2.62	0.34	0.89	0.91	4.39
			2	12	4.48	-0.11	1.94	0.91	7.70	0.68	1.80	0.90	12.16
		1	0.5	13	1.65	0.10	0.42	0.94	1.68	0.24	0.65	0.93	2.81
			1	14	2.72	0.10	1.01	0.93	4.06	0.52	1.27	0.92	6.41
			2	15	7.39	-0.97	3.73	0.85	13.69	1.18	3.19	0.90	21.49
		2	0.5	16	2.12	0.10	0.71	0.93	2.84	0.30	0.76	0.94	3.99
			1	17	4.48	-0.24	2.08	0.89	8.01	0.68	1.64	0.93	11.27
			2	18	20.09	-7.53	12.29	0.68	34.08	1.75	5.56	0.91	56.15
5	10	0.5	0.5	19	1.45	0.10	0.35	0.94	1.43	0.09	0.34	0.93	1.63
			1	20	2.12	0.14	0.75	0.93	2.96	0.17	0.55	0.92	3.32
			2	21	4.48	-0.11	2.18	0.89	8.12	0.30	0.96	0.90	9.26
		1	0.5	22	1.65	0.13	0.54	0.93	2.13	0.10	0.37	0.93	1.97
			1	23	2.72	0.12	1.25	0.91	4.74	0.20	0.61	0.93	4.52
			2	24	7.39	-1.10	4.40	0.80	14.55	0.37	1.10	0.91	15.59
		2	0.5	25	2.12	0.13	0.93	0.92	3.58	0.11	0.38	0.94	2.62
			1	26	4.48	-0.34	2.60	0.83	8.99	0.22	0.64	0.93	7.55
			2	27	20.09	-8.54	14.25	0.58	33.74	0.41	1.18	0.91	41.35
	20	0.5	0.5	28	1.45	0.05	0.23	0.95	0.94	0.09	0.32	0.94	1.58
			1	29	2.12	0.10	0.52	0.94	2.08	0.17	0.52	0.93	3.28
			2	30	4.48	0.05	1.62	0.92	6.48	0.34	0.93	0.93	9.61
		1	0.5	31	1.65	0.08	0.36	0.94	1.46	0.10	0.35	0.94	1.94
			1	32	2.72	0.11	0.89	0.93	3.57	0.20	0.57	0.94	4.50
			2	33	7.39	-0.52	3.34	0.88	12.88	0.40	1.04	0.93	16.47
		2	0.5	34	2.12	0.10	0.65	0.93	2.63	0.10	0.36	0.94	2.57
			1	35	4.48	-0.15	1.93	0.89	7.57	0.21	0.60	0.94	7.55
			2	36	20.09	-6.22	11.39	0.73	34.99	0.44	1.11	0.93	43.49

Table D.3: Frequentist properties of the estimators $\hat{\theta}_m^{B,GIG}$ and $\hat{\theta}_c^{B,GIG}(\nu_j)$ obtained with GIG priors having $\lambda = 0.5$.

						Global predictor				Conditioned Predictor			
n_g	m	ϕ	σ^2	Scen.	θ_m	Bias	RMSE	Cov.	Wid.	RABias	RRMSE	Aco	AWi
2	10	0.5	0.5	1	1.45	0.52	0.76	0.95	3.19	0.26	0.58	0.97	3.22
			1	2	2.12	0.90	1.53	0.96	6.41	0.43	0.97	0.96	6.45
			2	3	4.48	1.31	3.84	0.96	15.90	0.63	1.67	0.93	16.55
		1	0.5	4	1.65	0.61	1.00	0.95	4.11	0.33	0.70	0.96	3.88
			1	5	2.72	1.02	2.18	0.96	8.83	0.59	1.28	0.95	8.57
			2	6	7.39	0.70	6.46	0.91	24.55	1.01	2.64	0.92	26.43
		2	0.5	7	2.12	0.75	1.52	0.95	6.09	0.43	0.87	0.96	5.27
			1	8	4.48	0.92	3.88	0.93	14.82	0.82	1.78	0.94	14.23
			2	9	20.09	-5.63	17.03	0.71	50.40	1.69	4.86	0.91	65.21
	20	0.5	0.5	10	1.45	0.24	0.40	0.96	1.63	0.20	0.51	0.96	2.64
			1	11	2.12	0.51	0.93	0.96	3.75	0.38	0.89	0.95	5.58
			2	12	4.48	1.04	2.74	0.96	11.51	0.72	1.72	0.94	16.06
		1	0.5	13	1.65	0.31	0.56	0.96	2.26	0.27	0.63	0.95	3.27
			1	14	2.72	0.64	1.40	0.96	5.70	0.54	1.19	0.95	7.70
			2	15	7.39	0.83	4.81	0.94	20.02	1.14	2.71	0.93	26.80
		2	0.5	16	2.12	0.42	0.93	0.95	3.76	0.34	0.76	0.95	4.53
			1	17	4.48	0.68	2.67	0.95	11.02	0.73	1.57	0.94	13.14
			2	18	20.09	-3.76	13.42	0.82	48.83	1.74	4.42	0.93	67.56
5	10	0.5	0.5	19	1.45	0.31	0.51	0.96	2.22	0.11	0.34	0.95	1.86
			1	20	2.12	0.60	1.09	0.96	4.60	0.20	0.55	0.95	3.89
			2	21	4.48	1.10	3.05	0.96	12.64	0.34	0.92	0.94	11.18
		1	0.5	22	1.65	0.42	0.75	0.96	3.16	0.12	0.37	0.95	2.18
			1	23	2.72	0.74	1.71	0.95	7.02	0.23	0.61	0.94	5.10
			2	24	7.39	0.71	5.61	0.92	21.62	0.41	1.08	0.93	18.18
		2	0.5	25	2.12	0.54	1.24	0.95	5.10	0.13	0.39	0.95	2.85
			1	26	4.48	0.64	3.31	0.93	12.80	0.26	0.65	0.94	8.38
			2	27	20.09	-5.10	15.80	0.73	48.37	0.48	1.21	0.93	47.16
	20	0.5	0.5	28	1.45	0.14	0.29	0.95	1.17	0.10	0.32	0.95	1.71
			1	29	2.12	0.32	0.66	0.95	2.68	0.19	0.52	0.95	3.60
			2	30	4.48	0.79	2.09	0.96	8.65	0.37	0.92	0.94	10.83
		1	0.5	31	1.65	0.21	0.45	0.95	1.81	0.11	0.35	0.95	2.04
			1	32	2.72	0.46	1.12	0.96	4.57	0.22	0.58	0.95	4.82
			2	33	7.39	0.76	4.08	0.93	17.05	0.44	1.06	0.94	18.12
		2	0.5	34	2.12	0.33	0.80	0.95	3.28	0.12	0.37	0.95	2.69
			1	35	4.48	0.52	2.33	0.94	9.72	0.24	0.62	0.95	8.01
			2	36	20.09	-3.19	12.24	0.84	46.25	0.49	1.16	0.94	47.24

Table D.4: Frequentist properties of the estimators $\hat{\theta}_m^{B,GIG}$ and $\hat{\theta}_c^{B,GIG}(\nu_j)$ obtained with GIG priors having $\lambda = 2$.

Appendix E

BayesLN: Bayesian Inference for Log-Normal Data

The package is available for installation from GitHub at: https://github.com/aldogardini/BayesLN.
To install and load it in R, a version greater than 3.5.0 is required and the following code could be used:

```
if("devtools"%in%installed.packages()[,1]){
   library(devtools)
}else{
   install.packages("devtools")
   library(devtools)
}
install_github("aldogardini/BayesLN")
```

Package 'BayesLN'

January 9, 2020

Title Bayesian Inference for Log-Normal Data

Version 0.1.1

Description Bayesian inference under log-normality assumption must be performed very carefully. In fact, under the common priors for the variance, useful quantities in the original data scale (like mean and quantiles) do not have posterior moments that are finite. This package allows to easily carry out a proper Bayesian inferential procedure by fixing a suitable distribution (the generalized inverse Gaussian) as prior for the variance. Functions to estimate several kind of means (unconditional, conditional and conditional under a mixed model) and quantiles (unconditional and conditional) are provided.

Depends R (>= 3.5.0)

Imports optimx, ghyp, fAsianOptions, coda, Rcpp (>= 0.12.17), MASS, lme4, data.table

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.0.2

Suggests knitr, rmarkdown, RcppArmadillo

VignetteBuilder knitr

LinkingTo Rcpp, RcppArmadillo

Date 2020-01-09

R topics documented:

`EPA09` *Chrysene concentration data*

Description

Vector of 8 observations of chrysene concentration (ppb) found in water samples.

Usage

`EPA09`

Format

Numeric vector.

Source

USEPA. *Statistical analysis of groundwater monitoring data at rcra facilities: Unifed guidance.* Technical report, Office of Resource Conservation and Recovery, Program Implementation and Information Division, U.S. Environmental Protection Agency, Washington, D.C. (2009).

`fatigue` *Low cycle fatigue data*

Description

Data frame of 22 observations in 2 variables

Usage

`fatigue`

Format

Dataframe with variables:

`stress`: stress factor.

`cycle`: number of test cycles.

Source

Upadhyay, S. K., and M. Peshwani. *Posterior analysis of lognormal regression models using the Gibbs sampler.* Statistical Papers 49.1 (2008): 59-85.

GH_MGF *GH Moment Generating Function*

Description

Function that implements the moment generating function of the Generalized Hyperbolyc (GH) distribution.

Usage

```
GH_MGF(r, mu = 0, delta, alpha, lambda, beta = 0)
```

Arguments

r	Coefficient of the MGF. Can be viewd also as the order of the log-GH moments.
mu	Location parameter, default set to 0.
delta	Concentration parameter, must be positive.
alpha	Tail parameter, must be positive and greater than the modulus of beta.
lambda	Shape parameter.
beta	Skewness parameter, default set to 0 (symmetric case).

Details

This function allows to evaluate the moment generating function of the GH distribution in the point r. It is defined only for points that are lower than the value of γ, that is defined as: $\gamma^2 = \alpha^2 - \beta^2$. For integer values of r, it could also be considered as the *r*-th raw moment of the log-GH distribution.

laminators *Laminators*

Description

Data frame of 39 observations in 2 variables.

Usage

```
laminators
```

Format

Dataframe with variables:

Worker: label of the measured worker.

log_Y: logarithm of the measured Styrene concentration.

Source

R. H. Lyles, L. L. Kupper, and S. M. Rappaport. *Assessing regulatory compliance of occupational exposures via the balanced one-way random effects ANOVA model* Journal of Agricultural, Biological, and Environmental Statistics (1997).

LN_hierarchical *Bayesian estimation of a log - normal hierarchical model*

Description

Function that estimates a log-normal linear mixed model with GIG priors on the variance components, in order to assure the existence of the posterior moments of key functionals in the original data scale like conditioned means or the posterior predictive distribution.

Usage

```
LN_hierarchical(
  formula_lme,
  data_lme,
  y_transf = TRUE,
  functional = c("Subject", "Marginal", "PostPredictive"),
  data_pred = NULL,
  order_moment = 2,
  nsamp = 10000,
  par_tau = NULL,
  par_sigma = NULL,
  inits = list(NULL),
  verbose = TRUE,
  burnin = 0.1 * nsamp,
  n_thin = 1
)
```

Arguments

formula_lme	A two-sided linear formula object describing both the fixed-effects and random-effects part of the model is required. For details see lmer.
data_lme	Optional data frame containing the variables named in formula_lme.
y_transf	Logical. If TRUE, the response variable is assumed already as log-transformed.
functional	Functionals of interest: "Subject" for subject-specific conditional mean, "Marginal" for the overall expectation and "PostPredictive" for the posterior predictive distribution.
data_pred	Data frame with the covariate patterns of interest for prediction. All the covariates present in the data_lme object must be included. If NULL the design matrix of the model is used.
order_moment	Order of the posterior moments that are required to be finite.
nsamp	Number of Monte Carlo iterations.
par_tau	List of vectors defining the triplets of hyperparaemters for each random effect variance (as many vectors as the number of specified random effects variances).
par_sigma	Vector containing the tiplet of hyperparameters for the prior of the data variance.
inits	List of object for initializing the chains. Objects with compatible dimensions must be named with beta, sigma2 and tau2.
verbose	Logical. If FALSE, the messages from the Gibbs sampler are not shown.
burnin	Number of iterations to consider as burn-in.
n_thin	Number of thinning observations.

Details

The function allows to estimate a log-normal linear mixed model through a Gibbs sampler. The model equation is specified as in `lmer` model and the target functionals to estimate need to be declared. A weakly informative prior setting is automatically assumed, always keeping the finiteness of the posterior moments of the target functionals.

Value

The output list provided is composed of three parts. The object `$par_prior` contains the parameters fixed for the variance components priors. The object `$samples` contains the posterior samples for all the paramters. They are returned as a `mcmc` object and they can be analysed trough the functions contained in the coda package in order to check for the convergence of the algorithm. Finally, in `$summaries` an overview of the posteriors of the model parameters and of the target functionals is provided.

Examples

```
## Not run: library(BayesLN)
# Load the dataset included in the package
data("laminators")
data_pred_new <- data.frame(Worker = unique(laminators$Worker))
Mod_est<-LN_hierarchical(formula_lme = log_Y~(1|Worker),
                         data_lme = laminators,
                         data_pred = data_pred_new,
                         functional = c("Subject","Marginal"),
                         order_moment = 2, nsamp = 50000, burnin = 10000)

## End(Not run)
```

LN_hier_existence	*Numerical evaluation of the log-normal conditioned means posterior moments*

Description

Function that evaluates the existence conditions for moments of useful quantities in the original data scale when a log-normal linear mixed model is estimated.

Usage

```
LN_hier_existence(X, Z, Xtilde, order_moment = 2, s = 1, m = NULL)
```

Arguments

`X`	Design matrix for fixed effects.
`Z`	Design matrix for random effects.
`Xtilde`	Covariate patterns used for the leverage computation.
`order_moment`	Order of the posterior moments required to be finite.
`s`	Number of variances of the random effects.
`m`	Vector of size `s` (if s>1) that indicates the dimensions of the random effect vectors.

Details

This function computes the existence conditions for the moments up to order fixed by `order_moment` of the log-normal linear mixed model specified by the design matrices `X` and `Z`. It considers the prediction based on multiple covariate patterns stored in the rows of the `Xtilde` matrix.

Value

Both the values of the factors determining the existence condition and the values of the gamma parameters for the different variance components are provided.

`LN_Mean` *Bayesian Estimate of the Log-normal Mean*

Description

This function produces a Bayesian estimate of the log-normal mean, assuming a GIG prior for the variance and an improper flat prior for the mean in the log scale.

Usage

```
LN_Mean(
  x,
  method = "weak_inf",
  x_transf = TRUE,
  CI = TRUE,
  alpha_CI = 0.05,
  type_CI = "two-sided",
  nrep = 1e+05
)
```

Arguments

`x`	Vector containing the sample.
`method`	String that indicates the prior setting to adopt. Choosing `"weak_inf"` a weakly informative prior setting is adopted, whereas selecting `"optimal"` the hyperparameters are aimed at minimizing the frequentist MSE.
`x_transf`	Logical. If `TRUE`, the `x` vector is assumed already log-transformed.
`CI`	Logical. With the default choice `TRUE`, the posterior credibility interval is computed.
`alpha_CI`	Level of alpha that determines the credibility (`1-alpha_CI`) of the posterior interval.
`type_CI`	String that indicates the type of interval to compute: `"two-sided"` (default), `"UCL"` (i.e. Upper Credible Limit) for upper one-sided intervals or `"LCL"` (i.e. Lower Credible Limit) for lower one-sided intervals.
`nrep`	Number of simulations for the computation of the credible intervals.

Details

Summarizing the posterior mean of the log-normal expectation might be delicate since several common priors used for the variance do not produces posteriors with finite moments. The proposal by Fabrizi and Trivisano (2012) of adopting a generalized inverse Gaussian (GIG) prior for the variance in the log scale σ^2 has been implemented. Moreover, they discussed how to specify the hyperparameters according to two different aims.

Firstly, a weakly informative prior allowed to produce posterior credible intervals with good frequentist properties, whereas a prior aimed at minimizing the point estimator MSE was proposed too. The choice between the two priors can be made through the argument `method`.

The point estimates are exact values, whereas the credible intervals are provided through simulations from the posterior distribution.

Value

The function returns a list which includes the prior and posterior parameters, the point estimate of the log-normal mean that consists in the mean of the posterior distribution of the functional $\exp\{\mu + \sigma^2/2\}$ and the posterior variance.

Source

Fabrizi, E., & Trivisano, C. *Bayesian estimation of log-normal means with finite quadratic expected loss*. Bayesian Analysis, 7(4), 975-996. (2012).

Examples

```
# Load data
data("NCBC")
# Optimal point estimator
LN_Mean(x = NCBC$al, x_transf = FALSE, method = "optimal", CI = FALSE)
# Weakly informative prior and interval estimation
LN_Mean(x = NCBC$al, x_transf = FALSE, type_CI = "UCL")
```

`LN_MeanReg` *Bayesian Estimate of the conditional Log-normal Mean*

Description

This function produces a bayesian estimate of the conditional log-normal mean assuming a GIG prior for the variance and an improper prior for the regression coefficients of the linear regression in the log scale.

Usage

```
LN_MeanReg(
  y,
  X,
  Xtilde,
  method = "weak_inf",
  y_transf = TRUE,
  h = NULL,
```

```
  CI = TRUE,
  alpha_CI = 0.05,
  type_CI = "two-sided",
  nrep = 1e+05
)
```

Arguments

y	Vector of observations of the response variable.
X	Design matrix.
Xtilde	Matrix of covariate patterns for which an estimate is required.
method	String that indicates the prior setting to adopt. Choosing "weak_inf" a weakly informative prior setting is adopted, whereas selecting "optimal" the hyperparameters are aimed at minimizing the frequentist MSE.
y_transf	Logical. If TRUE, the y vector is already assumed as log-transformed.
h	Leverage. With the default option NULL, the average leverage is used.
CI	Logical. With the default choice TRUE, the posterior credibility interval is computed.
alpha_CI	Level of alpha that determines the credibility (1-alpha_CI) of the posterior interval.
type_CI	String that indicates the type of interval to compute: "two-sided" (default), "UCL" (i.e. Upper Credible Limit) for upper one-sided intervals or "LCL" (i.e. Lower Credible Limit) for lower one-sided intervals.
nrep	Number of simulations.

Details

In this function the same procedure as LN_Mean is implemented allowing for the inclusion of covariates. Bayesian point and interval estimates for the response variabile in the original scale are provided considering the model: $log(y_i) = X\beta$.

Value

The function returns a list including the prior and posterior parameters, the point estimate of the log-normal mean conditioned with respect to the covariate points included in Xtilde. It consists of the mean of the posterior distribution for the functional $\exp\{\tilde{x}_i^T\beta + \sigma^2/2\}$ and the posterior variance.

Source

Fabrizi, E., & Trivisano, C. *Bayesian Conditional Mean Estimation in Log-Normal Linear Regression Models with Finite Quadratic Expected Loss.* Scandinavian Journal of Statistics, 43(4), 1064-1077. (2016).

Examples

```
library(BayesLN)
data("fatigue")

# Design matrices
Xtot <- cbind(1, log(fatigue$stress), log(fatigue$stress)^2)
```

```
X <- Xtot[-c(1,13,22),]
y <- fatigue$cycle[-c(1,13,22)]
Xtilde <- Xtot[c(1,13,22),]
#Estimation
LN_MeanReg(y = y,
           X = X, Xtilde = Xtilde,
           method = "weak_inf", y_transf = FALSE)
```

`LN_Quant` *Bayesian estimate of the log-normal quantiles*

Description

This function produces an estimate for the log-normal distribution quantile of fixed level `quant`.

Usage

```
LN_Quant(
  x,
  quant,
  method = "weak_inf",
  x_transf = TRUE,
  guess_s2 = NULL,
  CI = TRUE,
  alpha_CI = 0.05,
  type_CI = "two-sided",
  method_CI = "exact",
  rel_tol_CI = 1e-05,
  nrep_CI = 1e+06
)
```

Arguments

`x`	Vector of data used to estimate the quantile.
`quant`	Number between 0 and 1 that indicates the quantile of interest.
`method`	String that indicates the prior setting to adopt. Choosing `"weak_inf"` a weakly informative prior setting is adopted, whereas selecting `"optimal"` the hyperparameters are fixed trough a numerical optimization algorithm aimed at minimizing the frequentist MSE.
`x_transf`	Logical. If `TRUE`, the `x` vector is assumed already log-transformed.
`guess_s2`	Specification of a guess for the variance if available. If not, the sample estimate is used.
`CI`	Logical. With the default choice `TRUE`, the posterior credibility interval is computed.
`alpha_CI`	Level of alpha that determines the credibility (1-`alpha_CI`) of the posterior interval.

`type_CI`	String that indicates the type of interval to compute: `"two-sided"` (default), `"UCL"` (i.e. Upper Credible Limit) for upper one-sided intervals or `"LCL"` (i.e. Lower Credible Limit) for lower one-sided intervals.
`method_CI`	String that indicates if the limits should be computed through the logSMNG quantile function `qlSMNG` (option `"exact"`, default), or by randomly generating a sample (`"simulation"`) using the function `rlSMNG`.
`rel_tol_CI`	Level of relative tolerance required for the `integrate` procedure or for the infinite sum. Default set to `1e-5`.
`nrep_CI`	Number of simulations in case of `method="simulation"`.

Details

The function allows to carry out Bayesian inference for the unconditional quantiles of a sample that is assumed log-normally distributed.

A generalized inverse Gaussian prior is assumed for the variance in the log scale σ^2, whereas a flat improper prior is assumed for the mean in the log scale ξ.

Two alternative hyperparamters setting are implemented (choice controlled by the argument `method`): a weakly informative proposal and an optimal one.

Value

The function returns the prior parameters and their posterior values, summary statistics of the log-scale parameters and the estimate of the specified quantile: the posterior mean and variance are provided by default. Moreover, the user can control the computation of posterior intervals.

Examples

```
library(BayesLN)
data("EPA09")
LN_Quant(x = EPA09, quant = 0.95, method = "optimal", CI = FALSE)
LN_Quant(x = EPA09, quant = 0.95, method = "weak_inf",
        alpha_CI = 0.05, type_CI = "UCL")
```

`LN_QuantReg` *Bayesian estimate of the log-normal conditioned quantiles*

Description

This function produces a point estimate for the log-normal distribution quantile of fixed level `quant`.

Usage

```
LN_QuantReg(
  y,
  X,
  Xtilde,
  quant,
  method = "weak_inf",
  guess_s2 = NULL,
```

```
  y_transf = TRUE,
  CI = TRUE,
  method_CI = "exact",
  alpha_CI = 0.05,
  type_CI = "two-sided",
  rel_tol_CI = 1e-05,
  nrep = 1e+05
)
```

Arguments

`y`	Vector of observations of the response variable.
`X`	Design matrix.
`Xtilde`	Covariate patterns of the units to estimate.
`quant`	Number between 0 and 1 that indicates the quantile of interest.
`method`	String that indicates the prior setting to adopt. Choosing `"weak_inf"` a weakly informative prior setting is adopted, whereas selecting `"optimal"` the hyperparameters are fixed trough a numerical optimization algorithm aimed at minimizing the frequentist MSE.
`guess_s2`	Specification of a guess for the variance if available. If not, the sample estimate is used.
`y_transf`	Logical. If `TRUE`, the `y` vector is assumed already log-transformed.
`CI`	Logical. With the default choice `TRUE`, the posterior credibility interval is computed.
`method_CI`	String that indicates if the limits should be computed through the logSMNG quantile function `qlSMNG` (option `"exact"`, default), or by randomly generating (`"simulation"`) using the function `rlSMNG`.
`alpha_CI`	Level of credibility of the posterior interval.
`type_CI`	String that indicates the type of interval to compute: `"two-sided"` (default), `"UCL"` (i.e. Upper Credible Limit) for upper one-sided intervals or `"LCL"` (i.e. Lower Credible Limit) for lower one-sided intervals.
`rel_tol_CI`	Level of relative tolerance required for the `integrate` procedure or for the infinite sum. Default set to `1e-5`.
`nrep`	Number of simulations for the C.I. in case of `method="simulation"` and for the posterior of the coefficients vector.

Details

The function allows to carry out Bayesian inference for the conditional quantiles of a sample that is assumed log-normally distributed. The design matrix containing the covariate patterns of the sampled units is `X`, whereas `Xtilde` contains the covariate patterns of the unit to predict.

The classical log-normal linear mixed model is assumed and the quantiles are estimated as:

$$\theta_p(x) = exp(x^T\beta + \Phi^{-1}(p))$$

.

A generalized inverse Gaussian prior is assumed for the variance in the log scale σ^2, whereas a flat improper prior is assumed for the vector of coefficients β.

Two alternative hyperparamters setting are implemented (choice controlled by the argument `method`): a weakly informative proposal and an optimal one.

Value

The function returns the prior parameters and their posterior values, summary statistics of the parameters β and σ^2, and the estimate of the specified quantile: the posterior mean and variance are provided by default. Moreover the user can control the computation of posterior intervals.

`NCBC` *Naval Construction Battalion Center data*

Description

Data frame of 17 observations in 2 variables

Usage

`NCBC`

Format

Dataframe with 2 variables:

`al`: aluminium concentration measures.

`mn`: manganese concentration measures.

Source

Singh, Ashok K., Anita Singh, and Max Engelhardt. *The lognormal distribution in environmental applications.* Technology Support Center Issue Paper. (1997).

`ReadingTime` *Reading Times data*

Description

Data frame of 547 observations in 4 variables

Usage

`ReadingTime`

Format

Dataframe with variables:

`subj`: label indicating the subject.

`item`: label indicating the item read.

`so`: variable assuming value 1 (object relative condition) and -1 (subject relative condition).

`log_rt`: logarithm of the reading time measured.

Source

E. Gibson and H.-H. I. Wu. *Processing chinese relative clauses in context.* Language and Cognitive Processes, 28(1-2):125-155. (2008).

SMNGdistribution *SMNG and logSMNG Distributions*

Description

Density function, distribution function, quantile function and random generator for the SMNG distribution and the logSMNG. It requires the specification of a five prameters vector: mu, delta, gamma, lambda and beta.

Usage

```
dSMNG(
  x,
  mu = 0,
  delta,
  gamma,
  lambda,
  beta = 0,
  inf_sum = FALSE,
  rel_tol = 1e-05
)

pSMNG(q, mu, delta, gamma, lambda, beta, rel_tol = 1e-05)

qSMNG(p, mu, delta, gamma, lambda, beta, rel_tol = 1e-05)

rSMNG(n, mu, delta, gamma, lambda, beta)

dlSMNG(x, mu = 0, delta, gamma, lambda, beta, inf_sum = FALSE, rel_tol = 1e-05)

plSMNG(q, mu, delta, gamma, lambda, beta, rel_tol = 1e-05)

qlSMNG(p, mu, delta, gamma, lambda, beta, rel_tol = 1e-05)

rlSMNG(n, mu, delta, gamma, lambda, beta)
```

Arguments

x, q	Vector of quantiles.
mu	Location parameter, default set to 0.
delta	Concentration parameter, must be positive.
gamma	Tail parameter, must be positive.
lambda	Shape parameter.
beta	Skewness parameter, default set to 0 (symmetric case).
inf_sum	Logical: if FALSE (default) the integral representation of the SMNG density is used, otherwise the infinite sum is employed.
rel_tol	Level of relative tolerance required for the integrate procedure or for the infinite sum convergence check. Default set to 1e-5.
p	Vector of probabilities.
n	Sample size.

Details

The SMNG distribution is a normal scale-mean mixture distribution with a GIG as mixing distribution. The density can be expressed as an infinite sum of Bessel K functions and it is characterized by 5 parameters.

Moreover, if X is SMNG distributed, then $Z = exp(X)$ is distributed as a log-SMNG distribution.

Value

`dSMNG` and `dlSMNG` provide the values of the density function at a quantile `x` for, respectively a SMNG distribution and a log-SMNG.

`pSMNG` and `plSMNG` provide the cumulative distribution function at a quantile `q`.

`qSMNG` and `qlSMNG` provide the quantile corresponding to a probability level `p`.

`rSMNG` and `rlSMNG` generate `n` independent samples from the desired distribution.

Examples

```
## Not run:
### Plots of density and cumulative functions of the SMNG distribution
x<-seq(-10,10,length.out = 500)
par(mfrow=c(1,2))

plot(x,dSMNG(x = x,mu = 0,delta = 1,gamma = 1,lambda = 1,beta= 2),
    type="l",ylab="f(x)")
lines(x,dSMNG(x = x,mu = 0,delta = 1,gamma = 1,lambda = 1,beta= -2),col=2)
title("SMNG density function")

plot(x,pSMNG(q = x,mu = 0,delta = 1,gamma = 1,lambda = 1,beta= 2),
    type="l",ylab="F(x)")
lines(x,pSMNG(q = x,mu = 0,delta = 1,gamma = 1,lambda = 1,beta= -2),col=2)
title("SMNG cumulative function")


### Plots of density and cumulative functions of the logSMNG distribution
x<-seq(0,20,length.out = 500)
par(mfrow=c(1,2))

plot(x,dlSMNG(x = x,mu = 0,delta = 1,gamma = 1,lambda = 2,beta = 1),
    type="l",ylab="f(x)",ylim = c(0,1.5))
lines(x,dlSMNG(x = x,mu = 0,delta = 1,gamma = 1,lambda = 2,beta = -1),col=2)
title("logSMNG density function")

plot(x,plSMNG(q = x,mu = 0,delta = 1,gamma = 1,lambda = 2,beta = 1),
    type="l",ylab="F(x)",ylim = c(0,1))
lines(x,plSMNG(q = x,mu = 0,delta = 1,gamma = 1,lambda = 2,beta = -1),col=2)
title("logSMNG cumulative function")

## End(Not run)
```

SMNGmoments *SMNG Moments and Moment Generating Function*

Description

Functions that implement the mean, the generic moments (both raw and centered) and the moment generating function of the SMNG distribution.

Usage

```
SMNG_MGF(
  r,
  mu = 0,
  delta,
  gamma,
  lambda,
  beta = 0,
  inf_sum = FALSE,
  rel_tol = 1e-05
)

meanSMNG(mu, delta, gamma, lambda, beta)

SMNGmoment(j, mu, delta, gamma, lambda, beta, type = "central")
```

Arguments

r	Coefficient of the MGF. Can be viewed also as the order of the logSMNG moments.
mu	Location parameter, default set to 0.
delta	Concentration parameter, must be positive.
gamma	Tail parameter, must be positive.
lambda	Shape parameter.
beta	Skewness parameter, default set to 0 (symmetric case).
inf_sum	Logical: if FALSE (default), the integral representation of the SMNG density is used, otherwise the infinite sum is employed.
rel_tol	Level of relative tolerance required for the `integrate` procedure or for the infinite sum. Default set to `1e-5`.
j	Order of the moment.
type	String that indicate the kind of moment to comupute. Could be `"central"` (default) or `"raw"`.

Details

If the mean (i.e. the first order raw moment) of the SMNG distribution is required, then the function `meanSMNG` could be use.

On the other hand, to obtain the generic j-th moment both `"raw"` or `"centered"` around the mean, the function `momentSMNG` could be used.

Finally, to evaluate the Moment Generating Function (MGF) of the SMNG distribution in the point `r`, the function `SMNG_MGF` is provided. It is defined only for points that are lower then the parameter `gamma`, and for integer values of `r` it could also considered as the *r*-th raw moment of the logSMNG distribution. The last function is implemented both in the integral form, which uses the routine `integrate`, or in the infinite sum structure.

Examples

```
### Comparisons sample quantities vs true values
sample <- rSMNG(n=1000000,mu = 0,delta = 2,gamma = 2,lambda = 1,beta = 2)
mean(sample)
meanSMNG(mu = 0,delta = 2,gamma = 2,lambda = 1,beta = 2)

var(sample)
SMNGmoment(j = 2,mu = 0,delta = 2,gamma = 2,lambda = 1,beta = 2,type = "central")
SMNGmoment(j = 2,mu = 0,delta = 2,gamma = 2,lambda = 1,beta = 2,type = "raw")-
                        meanSMNG(mu = 0,delta = 2,gamma = 2,lambda = 1,beta = 2)^2

mean(exp(sample))
SMNG_MGF(r = 1,mu = 0,delta = 2,gamma = 2,lambda = 1,beta = 2)
```

Index

Bibliography

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1964.

G. E. Andrews, R. Askey, and R. Roy. Special functions, volume 71 of encyclopedia of mathematics and its applications, 1999.

K. Antonio, J. Beirlant, T. Hoedemakers, and R. Verlaak. Lognormal mixed models for reported claims reserves. *North American Actuarial Journal*, 10(1):30–48, 2006.

O. Arslan. Variance-mean mixture of the multivariate skew normal distribution. *Statistical Papers*, 56(2):353–378, 2015.

N. Balakrishnan and D. Mitra. Likelihood inference for lognormal data with left truncation and right censoring with an illustration. *Journal of Statistical Planning and Inference*, 141(11):3536–3553, 2011.

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2014.

O. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 353(1674):401–419, 1977.

O. Barndorff-Nielsen, J. Kent, and M. Sørensen. Normal variance-mean mixtures and z distributions. *International Statistical Review/Revue Internationale de Statistique*, pages 145–159, 1982.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

G. E. Battese, R. M. Harter, and W. A. Fuller. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401):28–36, 1988.

P. Berchialla, I. Baldi, V. Notaro, S. Barone-Monfrin, F. Bassi, and D. Gregori. Flexibility of bayesian generalized linear mixed models for oral health research. *Statistics in medicine*, 28(28):3509–3522, 2009.

E. Berg and H. Chandra. Small area prediction for a unit-level lognormal model. *Computational Statistics & Data Analysis*, 78:159–175, 2014.

E. Berg, H. Chandra, and R. Chambers. Small area estimation for lognormal data. *Analysis of Poverty Data by Small Area Estimation*, pages 279–298, 2016.

J. O. Berger and J. M. Bernardo. On the development of the reference prior method. *Bayesian statistics*, 4:35–60, 1992.

B. M. Bibby and M. Sørensen. Hyperbolic processes in finance. *Handbook of heavy tailed distributions in finance*, 1:211–248, 2003.

H. D. Bondell, B. J. Reich, and H. Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838, 2010.

G. E. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211–252, 1964.

W. Breymann and D. Lüthi. ghyp: A package on generalized hyperbolic distributions. *Manual for R Package ghyp*, 2013.

L. Broemeling. *Bayesian analysis of linear models*. Statistics, textbooks and monographs. M. Dekker, 1985.

W. H. Bullock and J. S. Ignacio. *A strategy for assessing and managing occupational exposures*. AIHA, 2006.

P.-C. Bürkner. Advanced bayesian multilevel modeling with the r package brms. *R Journal*, 10(1), 2018.

B. P. Carlin and T. A. Louis. *Bayesian methods for data analysis*. CRC Press, 2008.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

K. Chaloner. A bayesian approach to the estimation of variance components for the unbalanced one-way random model. *Technometrics*, 29(3):323–337, 1987.

H. Chandra and R. Chambers. Small area estimation under transformation to linearity. 2008.

H. Chandra and R. Chambers. Multipurpose weighting for small area estimation. *Journal of Official Statistics*, 25(3):379–395, 2009.

V. Chernozhukov, I. Fernández-Val, and A. Galichon. Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125, 2010.

S. Chib and B. P. Carlin. On mcmc sampling in hierarchical longitudinal models. *Statistics and Computing*, 9(1):17–26, 1999.

W. Cochran. Some difficulties in the statistical analysis of replicated experiments. *Emp. Journ. Expt. Agrie., vi*, 157, 1938.

P. Coursaget, B. Yvonnet, J. Chiron, W. Gilks, N. Day, C. Wang, et al. Scheduling of revaccination against hepatitis b virus. *The Lancet*, 337(8751):1180–1183, 1991.

E. L. Crow and K. Shimizu. *Lognormal distributions*. Marcel Dekker New York, 1987.

G. A. Dagne. Bayesian transformed models for small area estimation. *Test*, 10(2):375–391, 2001.

M. J. Daniels. A prior for the variance in hierarchical models. *Canadian Journal of Statistics*, 27(3):567–578, 1999.

G. S. Datta and M. Ghosh. Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics*, pages 1748–1770, 1991.

D. Eddelbuettel and C. Sanderson. Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational Statistics & Data Analysis*, 71:1054–1063, 2014.

D. Eddelbuettel, R. François, J. Allaire, K. Ushey, Q. Kou, N. Russel, J. Chambers, and D. Bates. Rcpp: Seamless r and c++ integration. *Journal of Statistical Software*, 40(8): 1–18, 2011.

J. F. England Jr, T. A. Cohn, B. A. Faber, J. R. Stedinger, W. O. Thomas Jr, A. G. Veilleux, J. E. Kiang, and R. R. Mason Jr. Guidelines for determining flood flow frequency - bulletin 17c. Technical report, US Geological Survey, 2018.

E. Fabrizi and C. Trivisano. Bayesian estimation of log-normal means with finite quadratic expected loss. *Bayesian Analysis*, 7(4):975–996, 2012.

E. Fabrizi and C. Trivisano. Bayesian conditional mean estimation in log-normal linear regression models with finite quadratic expected loss. *Scandinavian Journal of Statistics*, 43(4):1064–1077, 2016.

E. Fabrizi, M. R. Ferrante, and C. Trivisano. Bayesian small area estimation for skewed business survey variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(4):861–879, 2018.

S. Favaro, A. Lijoi, and I. Pruenster. On the stick-breaking representation of normalized inverse gaussian priors. *Biometrika*, 99(3):663–674, 2012.

C. Fernandez and M. F. Steel. Bayesian regression analysis with scale mixtures of normals. *Econometric Theory*, 16(1):80–101, 2000.

D. Finney. On the distribution of a variate whose logarithm is normally distributed. *Supplement to the Journal of the Royal Statistical Society*, 7(2):155–161, 1941.

R. A. Fisher and W. A. Mackenzie. Studies in crop variation. ii. the manurial response of different potato varieties. *The Journal of Agricultural Science*, 13(3):311–320, 1923.

B. Fitzenberger, R. A. Wilke, and X. Zhang. Implementing box–cox quantile regression. *Econometric Reviews*, 29(2):158–181, 2009.

J. Foster, J. Greer, and E. Thorbecke. A class of decomposable poverty measures. *Econometrica: journal of the econometric society*, pages 761–766, 1984.

S. Fruhwirth-Schnatter and H. Wagner. Bayesian variable selection for random intercept modeling of gaussian and non-gaussian data. In J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, editors, *Bayesian Statistics 9*, pages 165–185. Oxford University Press, Oxford, 2011.

A. Gelman. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534, 2006.

J. Geweke. Bayesian inference in econometric models using monte carlo integration. *Econometrica: Journal of the Econometric Society*, pages 1317–1339, 1989.

J. Ghosh, Y. Li, R. Mitra, et al. On the use of cauchy prior distributions for bayesian logistic regression. *Bayesian Analysis*, 13(2):359–383, 2018.

R. D. Gibbons, D. K. Bhaumik, and S. Aryal. *Statistical methods for groundwater monitoring*, volume 59. John Wiley & Sons, 2009.

E. Gibson and H.-H. I. Wu. Processing chinese relative clauses in context. *Language and Cognitive Processes*, 28(1-2):125–155, 2013.

W. Gilchrist. *Statistical modelling with quantile functions.* CRC Press, 2000.

I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products.* Academic press, 2014.

R. B. Gramacy. monomvn: Estimation for multivariate normal and student-t data with monotone missingness. *R package version*, pages 1–8, 2010.

J. E. Griffin and P. J. Brown. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188, 2010.

E. Gumbel. La probabilité des hypotheses. *Comptes Rendus de l'Académie des Sciences (Paris)*, 209:645–647, 1939.

G. J. Hahn and W. Q. Meeker. *Statistical intervals: a guide for practitioners*, volume 92. John Wiley & Sons, 2011.

K. Hamed and A. R. Rao. *Flood frequency analysis*. CRC press, 1999.

E. Hammerstein. *Generalized hyperbolic distributions: theory and applications to CDO pricing*. PhD thesis, PhD thesis, Universität Freiburg, 2010.

D. A. Harville and A. G. Zimmermann. The posterior distribution of the fixed and random effects in a mixed-effects linear model. *Journal of Statistical Computation and Simulation*, 54(1-3):211–229, 1996.

A. Hector, S. von Felten, Y. Hautier, M. Weilenmann, and H. Bruelheide. Effects of dominance and diversity on productivity along ellenberg's experimental water table gradients. *PloS one*, 7(9):e43358, 2012.

D. R. Helsel et al. *Nondetects and data analysis. Statistics for censored environmental data.* Wiley-Interscience, 2005.

B. M. Hill. Inference about variance components in the one-way model. *Journal of the American Statistical Association*, 60(311):806–825, 1965.

J. P. Hobert and G. Casella. The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473, 1996.

R. J. Hyndman and Y. Fan. Sample quantiles in statistical packages. *The American Statistician*, 50(4):361–365, 1996.

J. Jiang, P. Lahiri, S.-M. Wan, et al. A unified jackknife theory for empirical best prediction with m-estimation. *The Annals of Statistics*, 30(6):1782–1810, 2002.

B. Jorgensen. *Statistical properties of the generalized inverse Gaussian distribution*, volume 9. Springer Science & Business Media, 1982.

M. A. Karim and J. U. Chowdhury. A comparison of four distributions used in flood frequency analysis in bangladesh. *Hydrological Sciences Journal*, 40(1):55–66, 1995.

F. Karlberg et al. Population total prediction under a lognormal superpopulation model. *Metron*, 58(3/4):53–80, 2000.

R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.

R. Koenker and G. Bassett. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

K. Krishnamoorthy and T. Mathew. Assessing occupational exposure via the one-way random effects model with balanced data. *Journal of agricultural, biological, and environmental statistics*, 7(3):440, 2002.

K. Krishnamoorthy, A. Mallick, and T. Mathew. Inference for the lognormal mean and quantiles based on samples with left and right type i censoring. *Technometrics*, 53(1): 72–83, 2011.

K. Lange. *Numerical analysis for statisticians*. Springer Science & Business Media, 2010.

J. F. Lawless. *Statistical models and methods for lifetime data*. John Wiley & Sons, 2003.

E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

E. Limpert, W. A. Stahel, and M. Abbt. Log-normal distributions across the sciences: Keys and clues. *AIBS Bulletin*, 51(5):341–352, 2001.

N. T. Longford. Small-sample estimators of the quantiles of the normal, log-normal and pareto distributions. *Journal of Statistical Computation and Simulation*, 82(9):1383–1395, 2012.

R. H. Lyles, L. L. Kupper, and S. M. Rappaport. Assessing regulatory compliance of occupational exposures via the balanced one-way random effects anova model. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 64–86, 1997a.

R. H. Lyles, L. L. Kupper, and S. M. Rappaport. A lognormal distribution-based exposure assessment method for unbalanced data. *The Annals of occupational hygiene*, 41(1):63–76, 1997b.

M. Machado, B. Botero, J. López, F. Francés, A. Díez-Herrero, and G. Benito. Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrology and Earth System Sciences*, 19(6):2561, 2015.

D. McAlister. Xiii. the law of the geometric mean. *Proceedings of the Royal Society of London*, 29(196-199):367–376, 1879.

R. A. Meyers. *Complex systems in finance and econometrics*. Springer Science & Business Media, 2010.

S. P. Millard. *EnvStats, an R Package for Environmental Statistics*. Wiley Online Library, 2013.

I. Molina and N. Martin. Empirical best prediction under a nested error model with log transformation. *The Annals of Statistics*, 46(5):1961–1993, 2018.

I. Molina and J. Rao. Small area estimation of poverty indicators. *Canadian Journal of Statistics*, 38(3):369–385, 2010.

I. Molina, B. Nandram, and J. Rao. Small area estimation of general parameters with application to poverty indicators: a hierarchical bayes approach. *The Annals of Applied Statistics*, 8(2):852–885, 2014.

J. C. Nash et al. On best practice optimization methods in r. *Journal of Statistical Software*, 60(2):1–14, 2014.

R. Natarajan and R. E. Kass. Reference bayesian methods for generalized linear mixed models. *Journal of the American Statistical Association*, 95(449):227–237, 2000.

M. S. Paolella. *Intermediate probability: A computational approach.* John Wiley & Sons, 2007.

D. Pfeffermann et al. New important developments in small area estimation. *Statistical Science*, 28(1):40–68, 2013.

J. Pilz, P. Pluch, and G. Spöck. Bayesian kriging with lognormal data and uncertain variogram parameters. In *Geostatistics for Environmental Applications*, pages 51–62. Springer, 2005.

M. Plummer. *rjags: Bayesian Graphical Models using MCMC*, 2016. URL `https://CRAN.R-project.org/package=rjags`. R package version 4-6.

N. G. Polson, J. G. Scott, et al. On the half-cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4):887–902, 2012.

P. N. Price, A. V. Nero, and A. Gelman. Bayesian prediction of mean indoor radon concentrations for minnesota counties. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 1995.

I. Prosdocimi, T. Kjeldsen, and C. Svensson. Non-stationarity in annual and seasonal series of peak flow and precipitation in the uk. *Natural Hazards and Earth System Sciences*, 14 (5):1125–1144, 2014.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

J. N. Rao and I. Molina. *Small area estimation.* John Wiley & Sons, 2015.

S. Rappaport, R. Lyles, and L. Kupper. An exposure-assessment strategy accounting for within-and between-worker sources of variability. *The Annals of occupational hygiene*, 39 (4):469–495, 1995.

C. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation.* Springer Science & Business Media, 2007.

A. L. Rukhin. Improved estimation in lognormal models. *Journal of the American Statistical Association*, 81(396):1046–1049, 1986.

D. J. Scott, D. Würtz, C. Dong, and T. T. Tran. Moments of the generalized hyperbolic distribution. *Computational statistics*, 26(3):459–476, 2011.

T. Sen, S. Singh, and Y. M. Tripathi. Statistical inference for lognormal distribution with type-i progressive hybrid censored data. *American Journal of Mathematical and Management Sciences*, 38(1):70–95, 2019.

H. Shen, L. D. Brown, and H. Zhi. Efficient estimation of log-normal means with application to pharmacokinetic data. *Statistics in medicine*, 25(17):3023–3038, 2006.

R. S. Silva, H. F. Lopes, and H. S. Migon. The extended generalized inverse gaussian distribution for log-linear and stochastic volatility models. *Brazilian Journal of Probability and Statistics*, pages 67–91, 2006.

D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn. Openbugs user manual, version 3.0. 2. *MRC Biostatistics Unit, Cambridge*, 2007.

J. R. Stedinger. Fitting log normal distributions to hydrologic data. *Water Resources Research*, 16(3):481–490, 1980.

W. Strupczewski, V. Singh, and H. Mitosek. Non-stationary approach to at-site flood frequency modelling. iii. flood analysis of polish rivers. *Journal of Hydrology*, 248(1-4): 152–167, 2001.

D. Sun and P. L. Speckman. A note on nonexistence of posterior moments. *Canadian Journal of Statistics*, 33(4):591–601, 2005.

D. Sun, R. K. Tsutakawa, and Z. He. Propriety of posteriors with improper priors in hierarchical linear mixed models. *Statistica Sinica*, 11(1):77–96, 2001.

L. Thabane and M. S. Haq. Prediction from a normal model using a generalized inverse gaussian prior. *Statistical Papers*, 40(2):175–184, 1999.

G. C. Tiao and W. Tan. Bayesian analysis of random-effect models in the analysis of variance. i. posterior distribution of variance-components. *Biometrika*, 52(1/2):37–53, 1965.

USEPA. Statistical analysis of groundwater monitoring data at rcra facilities: Unified guidance. Technical report, Office of Resource Conservation and Recovery, Program Implementation and Information Division, U.S. Environmental Protection Agency, Washington, D.C., 2009.

G. J. Van Breukelen. Psychometric modeling of response speed and accuracy with mixed and conditional regression. *Psychometrika*, 70(2):359–376, 2005.

G. Villarini, J. A. Smith, F. Serinaldi, J. Bales, P. D. Bates, and W. F. Krajewski. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. *Advances in Water Resources*, 32(8):1255–1266, 2009.

R. M. Vogel and I. Wilson. Probability distribution of annual maximum, mean, and minimum streamflows in the united states. *Journal of hydrologic Engineering*, 1(2):69–76, 1996.

R. M. Vogel, C. Yaindl, and M. Walter. Nonstationarity: Flood magnification and recurrence reduction factors in the united states. *JAWRA Journal of the American Water Resources Association*, 47(3):464–474, 2011.

C. Williams. The use of logarithms in the interpretation of certain entomological problems. *Annals of Applied Biology*, 24(2):404–414, 1937.

A. Winkelbauer. Moments and absolute moments of the normal distribution. *arXiv preprint arXiv:1209.4340*, 2012.

K. Ye. Bayesian reference prior analysis on the ratio of variances for the balanced one-way random effect model. *Journal of Statistical Planning and Inference*, 41(3):267–280, 1994.

K. Yu and R. A. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54 (4):437–447, 2001.

K. Yu, Z. Lu, and J. Stander. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):331–350, 2003.

Y. Yu. On normal variance–mean mixtures. *Statistics & Probability Letters*, 121:45–50, 2017.

A. Zellner. Bayesian and non-bayesian analysis of the log-normal distribution and log-normal regression. *Journal of the American Statistical Association*, 66(334):327–330, 1971.

X.-H. Zhou. Estimation of the log-normal mean. *Statistics in Medicine*, 17(19):2251–2264, 1998.

T. Zimmermann and R. T. Münnich. Small area estimation with a lognormal mixed model under informative sampling. *Journal of Official Statistics*, 34(2):523–542, 2018.