Enrico Fabrizi, Carlo Trivisano

Bayes estimators of log-normal means with finite quadratic expected loss

Quaderni di Dipartimento

Serie Ricerche, n. 6
ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche "Paolo Fortunati"

Hierarchical Bayes Analysis of the Log-normal Distribution Under Quadratic Loss Function

Enrico Fabrizi
DISES, Università Cattolica, Piacenza

Carlo Trivisano
Dipartimento di Scienze Statistiche 'P. Fortunati',
Università di Bologna

Abstract

The log-normal distribution is a popular model in biostatistics as in many other fields of statistics. Bayesian inference on the mean and median of the distribution is problematic because, for many popular choices of the prior for variance (on the log-scale) parameter, the posterior distribution has no finite moments, leading to Bayes estimators with infinite expected loss for the most common choices of the loss function. In this paper we propose a generalized inverse Gaussian prior for the variance parameter, that leads to a log-generalized hyperbolic posterior, a distribution for which it is easy to calculate quantiles and moments, provided that they exist. We derive the constraints on the prior parameters that yields finite posterior moments of order r. For the quadratic and relative quadratic loss functions, we investigate the choice of prior parameters leading to Bayes estimators with optimal frequentist mean square error. For the estimation of the lognormal mean we show, using simulation, that the Bayes estimator under quadratic loss compares favorably in terms of frequentist mean square error to known estimators. The theory does not apply only to the mean or median estimation but to all parameters that may be written as the exponential of a linear combination of the distribution's two parameters that include the mode and all non central moments.

Keywords: Bayes estimators, generalized hyperbolic distribution, generalized inverse gamma distribution, Bessel functions.

1 Introduction

Suppose that a random variable X with mean ξ and variance σ^2 is normally distributed, such that $\exp(X) \sim LogN(\xi, \sigma^2)$. In this paper we consider the problem of Bayesian inference about functionals of (ξ, σ^2) of the form $\theta_{a,b} = \exp(a\xi + b\sigma^2)$ with $a, b \in \Re$ based on a random sample $(X_1, \ldots, X_n)$. We may obtain the mean, the median, the mode and various non-central moments of the log-normal distribution for different choices of a, b. More specifically $a = 1$ and $b = 0$ yields the median ($\theta_{1,0}$), $a = 1$ and $b = -1$ yields the mode ($\theta_{1,-1}$) and $a = 1$ and $b = .5$ yields the mean ($\theta_{1,0.5}$). This estimation problem is of practical relevance because, when analyzing positively skewed data, it is common practice to take the log transformation and assume the normality of the transformed data. With reference to biological sciences see, among others, the review of Limpert et al. (2001) and Gill (2004). The problem has also a long tradition in both the frequentist and the Bayesian literature.

With reference to the latter, an important starting point is Zellner (1971). He considers diffuse priors of the type $p(\xi, \sigma) \propto \sigma^{-1}$ and he obtains the following results for the log-normal median: *i)* $p(\theta_{1,0}|\sigma, data)$ is a log-normal distribution; and *ii)* $p(\theta_{1,0}|data)$ is a log-t distribution.

Summarizing the log-t distribution is challenging using popular loss functions, such as the quadratic, because moments of all orders do not exist. For the log-normal mean Zellner (1971) shows that $p(\theta_{1,0.5}|\sigma, data)$ is a log-normal distribution. To obtain $p(\theta_{1,0.5}|data)$ he suggests studying the joint posterior $p(log(\theta_{1,0.5}), \sigma|data)$, integrating out σ and then considering its exponential transformation. He notes that the integral over σ can be 'expressed in terms of modified Bessel functions' but that 'it is the case that the posterior mean of $\theta_{1,0.5}$ does not exists', which creates similar problems in summarizing the posterior distribution, (that is in obtaining 'Bayes estimators' of the parameter in question and easily interpretable measures of the information loss).

Most of Zellner's paper focuses on the inference conditional on σ. He notes that, within the class of estimators of the form $k \exp(\bar{X})$ with $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and where k a is constant, the estimator for $\theta_{1,0.5}$ with minimum mean square error (MSE) is given by $\breve{\theta}_{1,0.5} = \exp(\bar{X} + \sigma^2/2 - 3\sigma^2/2n)$. From a Bayesian point of view, this estimator may be justified as the minimizer of the posterior expected loss, provided that the relative quadratic loss function $L_{RQ} = [(\theta - \hat{\theta})/\theta]^2$ is adopted.

Another important reference is Rukhin (1986). Rukhin proposes the following 'generalized' prior:

$$p(\xi, \sigma) = p(\sigma) \propto \sigma^{-2\nu+n-2} \exp\big(-\sigma^2\big[\gamma^2/2 - 2\big(b - a^2/n\big)\big]\big), \tag{1}$$

with $\gamma^2 > 4(b - a^2/n)$. Assuming the relative quadratic loss function L_{RQ}, he obtains an estimator for $\theta_{a,b}$ of the form $\hat{\theta}_{a,b}^{Ru} = \exp(a\bar{X})g(Y)$ that is given by

$$\hat{\theta}_{a,b}^{Ru} = \exp(a\bar{X}) \left(\frac{\beta}{\gamma}\right)^{\nu} \frac{K_\nu(\beta Y)}{K_\nu(\gamma Y)}, \tag{2}$$

$\beta = \gamma^2 - 2c$, $c = b - 3a^2/(2n)$ and $K_\nu(\,)$ is the modified Bessel function of the third kind (the Bessel-K function from now on). For a general introduction to Bessel functions, see Abramowitz and Stegun (1968), chapters 9 and 10. To obtain the values for the hyperparameters ν, γ, Rukhin (1986) chooses to minimize the frequentist MSE of $\hat{\theta}_{a,b}^{Ru}$. As the $K_\nu(\,)$ are quite difficult to handle, Rukhin uses a 'small arguments' approximation to $\hat{\theta}_{a,b}^{Ru}$ to propose a value for ν and a 'large arguments' approximation to propose a value for γ. Rukhin does not recognize that, with a simple change of variable, the prior he proposes may be seen as the product of a flat prior over the real line for ξ and the following prior on σ^2

$$p(\sigma^2) \propto (\sigma^2)^{-\nu+n/2-3/2} \exp(-\sigma^2[\psi^2/2 - 2\big(b - a^2/n)]) \tag{3}$$

which is the limit of a generalized inverse gamma distribution, $GIG(\lambda, \delta, \gamma)$ as $\delta \to 0$. The other parameters are given by $\lambda = -\nu + n/2 - 1/2$ and $\gamma^2 = \psi^2/2 - 2(b - a^2/n)$ (see section 2 for more details and notation). He does not provide the posterior distribution, so his proposal is inadequate for many inferential purposes (i.e., calculating of posterior variances or posterior probability intervals).

In this paper, we derive the posterior distribution of $\theta_{a,b}$ assuming a proper generalized inverse gamma prior on σ^2 (and a flat prior over the real line for ξ). We show that this posterior is a log-generalized hyperbolic distribution and state the conditions on the hyperparameters that guarantee the existence of posterior moments of a given order. Once these conditions are met for the first two non-central moments, we discuss the Bayes estimators with the ordinary quadratic loss function $L_Q = (\theta - \hat{\theta})^2$.

The main results of the paper may be summarized as follows: *i*) we show that, given our choice of the prior distributions, Bayes estimators associated with the relative quadratic loss function L_{RQ} can be reconducted to posterior expectations provided that b is properly modified; *ii*) adopting a 'small arguments' approximation to the Bessel-K functions and a choice of hyparameters aimed at minimizing the MSE, we show using simulation that our Bayes estimator of the mean, i.e. $\theta_{1,0.5}$ is substantially equivalent to the estimator proposed in Shen et al. (2006), which has been proven to be superior to many of the alternatives previously proposed in the literature.

The paper is organized as follows. In Section 2 we briefly present the generalized inverse Gaussian and generalized hyperbolic distributions. In Section 3, posterior distributions for σ^2 and $\theta_{a,b}$ are derived, and Bayes estimators under quadratic and relative quadratic losses are introduced. Section 4 is devoted to the choice of values to be assigned to the hyperparameters in order to obtain Bayes estimators with the minimum frequentist MSE. In Section 5 we introduce a simulation exercise and discuss the results, and Section 6 offers some conclusions and ideas for future research.

2 The generalized inverse Gaussian and generalized hyperbolic distributions

In this section we briefly introduce the generalized inverse Gaussian (GIG) and generalized hyperbolic (GH) distributions, establish the notation and mention some key properties that will be used later. For more details on these distributions, see Bibby and Sørensen (2003) and Eberlein and von Hammerstein (2004) among others.

The density of the GIG distribution may be written as follows:

$$p(x) = \Big(\frac{\gamma}{\delta}\Big)^{\lambda} \frac{1}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\Big\{ -\frac{1}{2}\big(\delta^2 x^{-1} + \gamma^2 x\big)\Big\} \mathbf{1}_{\Re^+} \tag{4}$$

If $\delta > 0$ the permissible values for the other parameters are $\gamma \geq 0$ if $\lambda < 0$ and $\gamma > 0$ if $\lambda = 0$. If $\delta \geq 0$ then γ, λ should be strictly positive.

The moments of the GIG can be expressed as functions of the Bessel-K functions by

$$E(X^j) = \left(\frac{\delta}{\gamma}\right)^j \frac{K_{\lambda+j}(\delta\gamma)}{K_\lambda(\delta\gamma)}. \tag{5}$$

The mode of the GIG is a simple function of the distribution's parameters. More specifically, $Mo(X) = (\gamma^{-2})\Big(\lambda - 1 + \sqrt{(\lambda-1)^2 + \delta^2\gamma^2}\Big)$ for strictly positive γ, and $Mo(X) = \delta^2/[2(1-\lambda)]$ for $\gamma = 0$. We note that using theorem 1.2 from Laforgia and Natalini (2010), and the positive skewness of the GIG distribution (Nguyen et al., 2003), it may easily be shown that for $\gamma > 0$

$$\frac{(\lambda-1) + \sqrt{(\lambda-1)^2 + \delta^2\gamma^2}}{\gamma^2} \leq E(X) \leq \frac{(\lambda+1) + \sqrt{(\lambda+1)^2 + \delta^2\gamma^2}}{\gamma^2}. \tag{6}$$

Many important distributions may be obtained as special cases of the GIG. For $\lambda > 0$ and $\gamma > 0$, the gamma distribution emerges as the limit when $\delta \to 0$. The inverse-gamma is obtained when $\lambda < 0$, $\delta > 0$ and $\gamma \to 0$ and an inverse Gaussian distribution is obtained when $\lambda = -\frac{1}{2}$.

Barndorff-Nielsen (1977) introduces the generalized hyperbolic (GH) distribution as a normal variance-mean mixture where the mixing distribution is GIG. That is, if $(X|W = w) \sim N(\mu + \beta w, w)$ and $W \sim GIG(\lambda, \delta, \gamma)$ then the marginal distribution of X will be GH (i.e., $X \sim GH(\lambda, \alpha, \beta, \delta, \mu)$, where $\alpha^2 = \beta^2 + \gamma^2$). The probability density function of the GH is given by

$$f(x) = \frac{\big(\frac{\gamma}{\delta}\big)^\lambda}{\sqrt{2\pi}K_\lambda(\delta\gamma)} \frac{K_{\lambda-1/2}\big(\alpha\sqrt{\delta^2 + (x-\mu)^2}\big)}{\big(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\big)^{1/2-\lambda}} \exp\big(\beta(x-\mu)\big)\mathbf{1}_{\Re}, \tag{7}$$

where $\gamma^2 = \alpha^2 - \beta^2$. The parameter domain is defined by the following conditions: *i)* $\delta \geq 0$, $\alpha > 0$, $\alpha^2 > \beta^2$ if $\lambda > 0$; *ii)* $\delta > 0$, $\alpha > 0$, $\alpha^2 > \beta^2$ if $\lambda = 0$; *iii)* $\delta > 0$, $\alpha \geq 0$, $\alpha^2 \geq \beta^2$ if $\lambda < 0$. The parameter α determines the shape, β determines the skewness (the sign of the skewness is consistent with that of β), μ is a location parameter, δ serves for scaling and λ influences the size of mass contained in the tails. The class of GH distributions is closed under affine transformations i.e if $X \sim GH(\lambda, \alpha, \beta, \delta, \mu)$ and $Z = b_0 X + b_1$ then $Z \sim GH(\lambda, \alpha/|b_0|, \beta/|b_0|, |b_0|\delta, b_0\mu + b_1)$. An essential tool in what follows is the moment generating function of the GH distribution:

$$M_{GH}(t) = \exp(\mu t)\left(\frac{\gamma^2}{\alpha^2-(\beta+t)^2}\right)^{\lambda/2}\frac{K_\lambda\big(\delta\sqrt{\alpha^2-(\beta+t)^2}\big)}{K_\lambda(\delta\gamma)} \tag{8}$$

which exists provided that $|\beta + t| < \alpha$.

3 Bayes estimators of $\theta_{a,b}$

3.1 Derivation of the posterior distribution for σ^2 and $\theta_{a,b}$

The representation of the GH distribution as a normal mean-variance mixture with the GIG as mixing distribution introduced in previous section provides the basis for obtaining the posterior distribution of $\eta_{a,b} = log(\theta_{a,b})$ when assuming a GIG prior for σ^2. More specifically we can prove the following result.

Theorem 3.1. *Assume the following: i) $p(\eta_{a,b}|\sigma^2, X) \sim N(\eta_{a,b}, a^2\sigma^2/n)$ and ii) $p(\xi,\sigma^2) = p(\xi)p(\sigma^2)$, with $p(\sigma^2) \sim GIG(\lambda,\delta,\gamma)$ and $p(\xi)$ an improper distribution uniform over the real line.*
It follows that

$$p(\sigma^2|data) \sim GIG(\bar\lambda, \bar\delta^\star, \gamma), \tag{9}$$

$$p(\eta_{a,b}|data) \sim GH(\bar\lambda, \bar\alpha, \bar\beta, \bar\delta, \bar\mu) \tag{10}$$

where $\bar\delta^\star = \sqrt{Y^2+\delta^2}$, $Y^2 = \sum_{i=1}^n (X_i - \bar X)^2$, $\bar\lambda = \lambda - \frac{n-1}{2}$, $\bar\alpha = \sqrt{\frac{n}{a^2}(\gamma^2 + \frac{nb^2}{a^2})}$ *and* $\bar\beta = n\frac{b}{a^2}$. *Let* $\bar\gamma^2 = \bar\alpha^2 - \bar\beta^2$. *As a consequence* $\bar\gamma^2 = \frac{n}{a^2}\gamma^2$, $\bar\delta = \sqrt{\frac{a^2}{n}(Y^2+\delta^2)}$ *and* $\bar\mu = a\bar X$.

Proof. To prove (9) simply note that:

$$\begin{aligned}
p(\sigma^2|data) &\propto \int \sigma^{-n}\exp\Big\{-\frac{1}{2\sigma^2}(Y^2 + n(\xi-\bar X)^2)\Big\} \\
&\times \Big(\frac{\gamma}{\delta}\Big)^\lambda \frac{1}{2K_\lambda(\delta\gamma)}(\sigma^2)^{\lambda-1}\exp\Big\{-\frac{1}{2}\big(\delta^2\sigma^{-2}+\gamma^2\sigma^2\big)\Big\}d\xi \\
&\propto \int (\sigma^2)^{-n/2+\lambda-1}\exp\Big\{-\frac{1}{2\sigma^2}\big(Y^2+\delta^2\big)-\frac{\gamma^2\sigma^2}{2}\Big\} \\
&\times \exp\Big\{-\frac{1}{2\sigma^2}n(\xi-\bar X)^2\Big\}d\xi \\
&= (\sigma^2)^{\lambda-\frac{n+1}{2}}\exp-\Bigg\{\frac{1}{2}\Big(\frac{Y^2+\delta^2}{\sigma^2}-\frac{\gamma^2\sigma^2}{2}\Big)\Bigg\}.
\end{aligned}$$

The second statement is a special case of Barndorff-Nielsen (1977) result. □

We are not primarily interested in $p(\eta_{a,b}|data)$, but rather in $\theta_{a,b} = \exp(\eta_{a,b})$ which is distributed as a log-GH, a distribution that has not, to our knowledge, received any attention in the literature. In any case, we can calculate the moments of $p(\theta_{a,b}|data)$ that we need for summarizing the posterior distribution with a quadratic loss function by using the moment-generating function of the GH distribution ($M_{GH}(t)$) and more specifically the fact that $E(\theta|data) = M_{GH}(1)$ and $V(\theta|data) = M_{GH}(2) - [M_{GH}(1)]^2$.

If we are able to generate samples from the GH distribution, moreover, we may obtain a sample from its exponential transformation. The quantiles and probability intervals may then be calculated using MC techniques. From among the variety of software available for generating random GH numbers, we mention the `ghyp` package running under `R` (Breymann and Lüthi, 2010).

$M_{\eta|data}(t)$ exists only if $\left|\bar{\beta} + t\right| < \bar{\alpha}$, or equivalently if $\bar{\alpha}^2 - (\bar{\beta} + t)^2 > 0$ (i.e., $\bar{\gamma}^2 > t^2 + 2n\frac{b}{a^2}t$). This condition implies the following constraint on the prior parameter γ:

$$\gamma^2 > \frac{a^2}{n}t^2 + 2bt. \tag{11}$$

The existence of posterior moments requires that γ is above a positive threshold when $a \neq 0, b > 0$ (as for the expected value). The threshold is asymptotically 0 for the median, (i.e., $\theta_{0,1}$), and it is negative for the mode (whenever $n > t/2$): so, it does not represent a restriction. With respect to the inference on the expected value ($\theta_{1,0.5}$) note that the popular inverse gamma prior on σ^2, a special case of the GIG for $\lambda < 0$, $\delta > 0$ when $\gamma \to 0$, does not respect condition (11) thereby leading to a posterior distribution with non-existent moments. Note that this result is consistent with the following remark from Zellner (1971) concerning the inference about $\theta_{1,0}$: posterior moments exist only for the limit as $n \to \infty$ (that is, when the log-t posterior converges to the log-normal).
Similarly, the uniform prior over the range $(0, A)$ for σ (Gelman , 2006) implies that $p(\sigma^2) \propto \frac{1}{\sigma}\mathbf{1}_{(0,A)}$, which may be seen as an approximation to a $Gamma(\frac{1}{2}, \epsilon)$ (where $\epsilon = (4A^2)^{-1}$) truncated at A^2. For $\lambda > 0$, $\gamma > 0$ and $\delta \to 0$, $GIG(\lambda, \delta, \gamma) \to Gamma(\lambda, \gamma^2/2)$. If we let $A \to \infty$, therefore, $p(\sigma) \propto 1$ is equivalent to a GIG prior with $\gamma \to 0$ and thus implies non-existent posterior moments.

Consistent with intuition, condition (11) implies that, in practice, to obtain a posterior distribution of $\theta_{a,b}$ with finite moments, a prior with short tails should be chosen.

3.2 Bayes estimators under quadratic and relative quadratic losses

If we summarize $p(\theta|data)$ using the ordinary quadratic loss function we obtain $\hat{\theta}^{QB}_{a,b} = E(\theta_{a,b}|data)$ or

$$\begin{aligned} \hat{\theta}^{QB}_{a,b} &= \exp(\bar{\mu})\left(\frac{\bar{\gamma}^2}{\bar{\alpha}^2-(\bar{\beta}+1)^2}\right)^{\bar{\lambda}/2} \\ &\times \frac{K_{\bar{\lambda}}\left(\bar{\delta}\sqrt{\bar{\alpha}^2-(\bar{\beta}+1)^2}\right)}{K_{\bar{\lambda}}(\bar{\delta}\bar{\gamma})} \qquad (12) \\ &= \exp(a\bar{X})\left(\frac{\gamma^2}{\gamma^2-(\frac{a^2}{n}+2b)}\right)^{(\lambda-\frac{n-1}{2})/2} \\ &\times \frac{K_{\{\lambda-\frac{n-1}{2}\}}\left(\sqrt{(Y^2+\delta^2)(\gamma^2-(\frac{a^2}{n}+2b))}\right)}{K_{\{\lambda-\frac{n-1}{2}\}}\left(\sqrt{(Y^2+\delta^2)\gamma^2}\right)}. \qquad (13) \end{aligned}$$

We provided two alternative expressions for $\hat{\theta}^{QB}_{a,b}$: (12) is indexed on the posterior parameters, and (13) highlights the role of the prior parameters, which and will be useful for studying the choice of hyperparameters that is discussed in the next section.

Under a relative quadratic loss function, the Bayes estimator is defined as $\hat{\theta}^{RQB}_{a,b} = E(\theta^{-1}_{a,b})/E(\theta^{-2}_{a,b})$ (see Zellner (1971)). The following result shows that $\hat{\theta}^{RQB}$ may be reconducted to a Bayes predictor under a quadratic loss function with a different choice of b and modified prior parameters.

Theorem 3.2. *For the Bayes estimator under relative quadratic loss function, we have that $\hat{\theta}^{RQB}_{a,b} = \hat{\theta}^{QB}_{a,b^\star}$ with $b^\star = b - 2a^2/n$ provided that the prior $p(\sigma^2) \sim GIG(\lambda, \delta, \gamma_\star)$ with $\gamma_\star^2 = \gamma^2 - 4a^2/n + 4b$ is assumed.*

Proof. Let $\tau_{a,b} = -log(\theta_{a,b})$. From the stated properties of the GIG distribution, we have $\tau_{a,b}|data \sim GH(\bar{\lambda}, \bar{\alpha}, \bar{\beta}, \bar{\delta}, -\bar{\mu})$ and
$2\tau_{a,b}|data \sim GH(\bar{\lambda}, \bar{\alpha}/2, \bar{\beta}/2, 2\bar{\delta}, -2\bar{\mu})$. Using (8) it may be shown that

$$\begin{aligned} \hat{\theta}^{RQB} &= \exp(a\bar{X})\left[\frac{\frac{n}{a^2}(\gamma^2-\frac{4a^2}{n}+4b)}{\frac{n}{a^2}(\gamma^2-\frac{a^2}{n}+2b)}\right]^{\{\lambda-\frac{n-1}{2}\}/2} \qquad (14) \\ &\times \frac{K_{\{\lambda-\frac{n-1}{2}\}}\left(\sqrt{(Y^2+\delta^2)(\gamma^2-\frac{a^2}{n}+2b)}\right)}{K_{\{\lambda-\frac{n-1}{2}\}}\left(\sqrt{(Y^2+\delta^2)(\gamma^2-\frac{4a^2}{n}+4b)}\right)} \end{aligned}$$

If we set $b_\star = b - 2a^2/n$ and $\gamma_\star^2 = \gamma^2 - \frac{4a^2}{n} + 4b$ we obtain a formula that has exactly the structure of (13).

□

4 Choice of hyperparameters

We can easily see that $\hat{\theta}^{QB}_{a,b}$ is sensitive to the choice of the prior parameters; therefore a careful choice of λ, δ, γ is an essential part of the inferential procedure. Following Rukhin (1986), our aim is to choose the hyperameters to minimize the frequentist MSE of the Bayes estimators. In practice this choice is a complicated task because expression (13) contains a ratio of Bessel-K functions that is quite intractable. Following Rukhin (1986) again, we will use a 'small argument' approximation to obtain the MSE-'optimal' values of the hyperparameters. Unfortunately this method is viable only for λ and δ because the small value approximation is free of γ. A more 'heuristic' argument will be proposed for the latter parameter. The simulation in Section 5 shows that the parameters determined in this manner also leads to good estimators performance when the arguments of the Bessel-K functions are no longer small. According to (11), priors with light tails are required to guarantee a finite posterior expectation and expected loss when $b > 0$. For this reason, when a guess of σ^2 is available a priori, it may be used to improve the performances of the Bayes estimators. How this can be done, without breaking down the MSE even when the guess is grossly wrong is described in section 4.2.

4.1 Choice of hyperparameters using the small arguments approximation to the modified Bessel functions of the third kind

Consider first the following approximation of $\hat{\theta}^{QB}_{a,b}$ using the 'small argument' approximation of the Bessel-K functions.

Theorem 4.1. *Under the assumptions that* $(Y^2+\delta^2)\gamma^2 < 1$, $(Y^2+\delta^2)(\gamma^2 - (\frac{a^2}{n}+2b)) < 1$ *and* $\lambda < \frac{n-1}{2}$ *we have that*

$$\hat{\theta}^{QB} \cong \exp(a\bar{X})\exp\left\{-\frac{(Y^2+\delta^2)(a^2+2nb)}{4n(\lambda-\frac{n-3}{2})}\right\} = \hat{\theta}^{qb} \qquad (15)$$

Proof. Now consider the following power series representation of a Bessel function of the first kind:

$$I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\Gamma(\nu+k+1)} \qquad (16)$$

$\nu \in \Re$. Note that for $z < 1$, the addends in the sum part of (16) are decreasing, so for small values ($z \to 0$) the series may be approximated by its first terms ($k = 0, 1, 2, \dots$). Moreover,

$$K_\nu(z) = \frac{\pi}{2\sin(\nu\pi)}\Big[I_{-\nu}(z) - I_\nu(z)\Big] \qquad (17)$$

$\nu \in \Re - Z$. Note that both (16) and (17) are proved in Abramowitz and Stegun (1968), chapter 9. Combining the two and assuming $\nu > 0$ we may write

$$\begin{aligned}\Big[I_{-\nu}(z) - I_{\nu}(z)\Big] &= \left(\frac{1}{2}z\right)^{-\nu}\Bigg[\frac{1}{\Gamma(1-\nu)} + \frac{\left(\frac{1}{4}z^2\right)}{\Gamma(2-\nu)} + \dots \\ &- \left(\frac{1}{2}z\right)^{2\nu}\frac{1}{\Gamma(\nu+1)} - \left(\frac{1}{2}z\right)^{2\nu+2}\frac{1}{\Gamma(\nu+2)} - \dots\Bigg]\end{aligned}$$

Note that the terms in the second line are negligible for $z \to 0$. Assuming $\nu - 1 \cong \nu$ and $(\nu-1)(\nu-2) \cong (\nu-1)^2$

$$K_{\nu}(z) \cong \frac{\pi}{2\sin(\nu\pi)}\left(\frac{1}{2}z\right)^{-\nu}\frac{1}{\Gamma(1-\nu)}\exp\left(\frac{1}{4}\frac{z^2}{1-\nu}\right). \tag{18}$$

We may re-write $\hat{\theta}^{QB}$ as

$$\hat{\theta}^{QB} = \exp(a\bar{X})\left(\frac{\gamma^2 - u}{\gamma^2}\right)^{\nu/2}\frac{K_{\nu}\left(\sqrt{(Y^2+\delta^2)(\gamma^2-u)}\right)}{K_{\nu}\left(\sqrt{(Y^2+\delta^2)\gamma^2}\right)},$$

where $u = \frac{a^2+2nb}{n}$, $\nu = -\bar{\lambda}$. Note that, as we assume $2\lambda < n-1$, the case $\nu < 0$ is not relevant. Using the fact that $K_a(\,) = K_{-a}(\,)$ and replacing z with the arguments of the Bessel-K that appear in $\hat{\theta}^{QB}$ complete the proof. □

Rukhin (1986) proves that to minimize the frequentist MSE of estimators in the form $\exp(aX)g(Y)$ such as $\hat{\theta}^{qb}$, and $\hat{\theta}^{QB}$,

$$E\Big[g(Y) - \exp(c\sigma^2)\Big]^2, \tag{19}$$

where $c = b - 3a^2/(2n)$, should be minimized. Unfortunately, minimization of (19) with respect to (λ, δ) does not lead to a unique minimum. The optimum MSE is reached for a set of (λ, δ) pairs that are described by equation (20).

Theorem 4.2. *Under the assumptions of theorem 4.1, the value of* λ *in (15) that minimizes (19) is given by*

$$\lambda_{opt} = \frac{n-3}{2} - \frac{(n-1)(a^2+2nb)}{4nc} - \frac{(a^2+2nb)}{4nc}\frac{\delta^2}{\sigma^2} \tag{20}$$

provided that $b \notin (-\frac{a^2}{2n}, \frac{3a^2}{2n})$. *The result holds for any* δ *in* $\Re^+$.

Proof. Following (19) we should minimize

$$\phi = E\left\{\exp\left[\frac{a^2+2nb}{2n(\nu-1)}(Y^2+\delta^2)\right] + \exp(c\sigma^2) - 2\exp\left[\frac{a^2+2nb}{4n(\nu-1)}(Y^2+\delta^2) + c\sigma^2\right]\right\}$$

where $\nu = -\bar{\lambda}$. Because $\frac{Y^2}{\sigma^2} \sim \chi^2(n-1) = Gamma\big(\frac{n-1}{2}, \frac{1}{2}\big)$ then $Y^2 \sim Gamma\big(\frac{n-1}{2}, \frac{1}{2\sigma^2}\big)$. Using the standard formula for the moment generating function it follows that $E\Big[\exp\Big(rY^2\Big)\Big] = \Big(1-2r\sigma^2\Big)^{-\frac{n-1}{2}}$, provided that $r < \frac{1}{2\sigma^2}$. Thus

$$E\Bigg\{\exp\Bigg[\frac{a^2+2nb}{2n(\nu-1)}Y^2\Bigg]\Bigg\} = \Bigg(1-\frac{a^2+2nb}{n(\nu-1)}\sigma^2\Bigg)^{-\frac{n-1}{2}} \cong \exp\Bigg(\frac{(n-1)(a^2+2nb)\sigma^2}{2n(\nu-1)}\Bigg)$$

when $\sigma^2 < \frac{n(\nu-1)}{a^2+2nb}$. The approximated function to be minimized may be written as

$$\tilde{\phi} = \exp\Bigg(\frac{(n-1)(a^2+2nb)\sigma^2}{2n(\nu-1)}\Bigg) - 2\exp\Bigg(\frac{(n-1)(a^2+2nb)\sigma^2}{4n(\nu-1)}\Bigg) + \exp(c\sigma^2)$$

Taking partial derivatives of $\tilde{\phi}$, and equating both partial derivatives to 0 leads to exactly the same equation:

$$\exp\Bigg\{\frac{(a^2+2nb)}{2n(\nu-1)}\Big[(n-1)\sigma^2+\delta^2\Big]\Bigg\} = \exp\Bigg\{\frac{(a^2+2nb)}{4n(\nu-1)}\Big[(n-1)\sigma^2+\delta^2\Big]+c\sigma^2\Bigg\}.$$

Taking the log and solving for ν, we obtain

$$\nu = 1 + \frac{(n-1)(a^2+2nb)}{4nc} + \frac{(a^2+2nb)}{4nc}\frac{\delta^2}{\sigma^2}. \tag{21}$$

So, we may obtain an 'optimal' value of ν for any choice of (positive) δ^2. Formula (20) follows from noting that $\nu = \frac{n-1}{2} - \lambda$. $\square$

Note that if $b \in (-\frac{a^2}{2n}, \frac{3a^2}{2n})$ we would obtain $\lambda_{opt} > \frac{n-1}{2}$ for which the approximation (15) on which this choice of λ is based is not valid anymore. In practice, the median ($a=1, b=0$) is the only relevant case that falls outside the applicability of theorem 4.2. To choose λ when $b=0$ we may observe that the estimator obtain under quadratic loss, (13), and relative quadratic loss, (14), are very close and asymptotically equivalent. Since $b^\star \notin (-\frac{a^2}{2n}, \frac{3a^2}{2n}$ we may then apply (20) replacing $b^\star = b - \frac{2a^2}{n}$ instead of b.

The λ_{opt} in (20) is a function not only of δ, as anticipated, but also of the unknown σ^2; therefore, an optimal choice of (δ, λ) should depend, at least in principle, on a prior guess for σ^2. A method for circumventing the problem that is implicitly suggested by the generalized prior (3), is to let $\delta \to 0$. This condition may be approximated in practice by a δ that is much smaller than σ so as to make the third addend in (20) negligible. This approximation can be justified by noting (from 6) that δ has the same order of magnitude of the expectation and the mode of σ^2; therefore, choices of the type $\delta = k\sigma_0^2$ for some constant k and prior guess of the variance σ_0^2, imply a negligible third addend in (20) in the 'small value setting' we assumed for the derivation of $\hat{\theta}^{qb}$.

The estimator $\hat{\theta}^{qb}_{1,0.5}$ is connected to popular estimators that have already been discussed in the literature. Note that if $\delta = 0$, $a = 1$, $b = 0.5$ and we replace λ_{opt} into (3.1), we obtain $\hat{\theta}^{qb}_{1,0.5} = \exp\big((\bar{X}) + \frac{S^2(n-3)}{2n}\big)$, which is the MSE-optimal estimator (3.9) proposed by Zellner (1971) with the assumed known σ^2 replaced by $S^2 = Y^2/(n-1)$. This estimator is also similar to the one proposed in Evans and Shaban (1976) with the function g truncated to the first term.

As far as γ is concerned, we propose choosing a value close to the minimum value to assure for the existence of the first two posterior moments. Therefore, we specify the GIG with heaviest possible tail among those yielding $p(\theta|data)$ with finite variance:

$$\gamma_0^2 = \max\Big\{0, 4\Big(\frac{a^2}{n} + b\Big)\Big\} + \epsilon. \tag{22}$$

where ϵ is a positive, typically small constant. Note that γ_0 depends on n. In any case, we found in our simulations that $\hat{\theta}^{QB}_{1,0.5}$ are not particularly sensitive to alternative choices of γ_0 that are close to (i.e., of the same order of magnitude of) the γ_0 we propose. Much larger values lead to inefficient $\hat{\theta}^{QB}_{1,0.5}$ with far larger frequentist MSEs.

Moreover, the resulting GIG distribution will have a particularly light right tail for positive b. This result implies that the prior will be relatively peaked. Because a choice δ close to 0 implies a peak close to 0, in the next section we explore how a prior guess on the population variance may be used to improve the performances of the Bayes estimators.

4.2 Choice of hyperarameters based on prior guesses of σ^2

Assume that a prior guess σ_0^2 for σ^2 is available and this value is 'far' from 0. The subject of this section is how to specify priors that incorporate this prior guess and preserve the optimality properties we have described. We restrict our attention to positive b, because when $b \leq 0$, the existence of the first posterior moment does not necessarily imply a prior with light tails, and we may choose priors with a more distributed probability mass.

We start by introducing the following approximation of $E(\sigma^2)$ and that is based on the inequality (6) and that has proved in simulations to be effective for a wide range of choices for λ, δ, γ:

$$E(\sigma^2) \cong \frac{\lambda + \sqrt{\lambda^2 + \delta^2\gamma^2}}{\gamma^2} \tag{23}$$

In the following we denote the approximate expected value from (23) by σ_E^2. We then introduce the following result:

Theorem 4.3. *Assuming that i) λ is a function of δ as expressed in (20) and that ii) $b > \frac{3a^2}{2n}$, we have $\sigma_E^2 = k\sigma^2$ for $0 < k < 1$.*

Proof. We may rewrite (20) as $\lambda = c_1 - c_2\delta^2$ with $c_1 = -1 - (n-1)\frac{2a^2}{2nb-3a^2}$ and $c_2 = \frac{a^2+2nb}{2(2nb-3a^2)\sigma^2}$. Substituting this identity into (23) and solving for δ^2, we obtain:

$$\delta^2 = \frac{k\sigma^2(k\gamma^2\sigma^2 - 2c_1)}{1 - 2kc_2\sigma^2}. \tag{24}$$

Of course, δ^2 must be positive. The signs of both c_1 and c_2 depends on that of c. Because if $2nb - 3a^2 > 0$ (i.e., $b > \frac{3a^2}{2n}$) implies that $c1 < 0$ and $c2 > 0$, the numerator of δ^2 is always positive and we may focus on the denominator. It is positive whenever

$$k < \frac{2nb - 3a^2}{2nb + a^2} < 1 \tag{25}$$

a condition that reduces to $k < \frac{n-3}{n+1}$ for $a = 1$, $b = 0.5$. □

The main implication of this result is that the relationship between λ and δ implied by (20) leads naturally to conservative priors whose expected values cannot exceed σ^2. This finding is consistent with all the theory of MSE-optimal estimation of log-normal parameters, in which efficiency is improved at the price of some negative bias. As a byproduct, (24) suggests how to choose δ^2 as a function of γ^2, a prior guess σ_0^2 and k.

To clarify the interpretation of the parameter k, we study its relation to the coefficient of variation of the prior distribution. Note first that if $W \sim GIG(\lambda, \delta, \gamma)$, with $\lambda < -1$, then

$$CV^2(W) \cong -\frac{1}{\lambda + 1}. \tag{26}$$

To see how this result follows, note that

$$CV^2(W) + 1 = \frac{K_{\lambda+2}(\delta\gamma)}{K_{\lambda+1}(\delta\gamma)} \frac{K_{\lambda}(\delta\gamma)}{K_{\lambda+1}(\delta\gamma)},$$

that is, $CV^2(W) + 1 = E(W_1)/E(W_2)$, with $W_1 \sim GIG(\lambda + 1, \delta, \gamma)$ and $W_2 \sim GIG(\lambda, \delta, \gamma)$. Using the approximation (23) for the expected values, we find that

$$CV^2(W) + 1 = \frac{(\lambda + 1) + \sqrt{(\lambda + 1)^2 + \delta^2\gamma^2}}{\lambda + \sqrt{\lambda^2 + \delta^2\gamma^2}}.$$

Because $\sqrt{m_1^2 + m_2} \cong |m_1| + \frac{m_2}{2|m_1|}$ for a positive m_2, $CV^2(W) + 1 \cong \lambda(\lambda + 1)^{-1}$ for $\lambda < -1$, from which (26) follows. We can now state the following result.

Theorem 4.4. *Given assumptions i) and ii) of theorem 4.3, assuming that δ is selected according to (24) and using the approximation in (26) for the squared coefficient of variation, if follows that $CV^2(\sigma^2)$ is a monotonically decreasing function of k.*

Proof. Because $b > \frac{3a^2}{2n}$, implies that $c_1 + 1 < 0$ and $c_2 > 0$, $\lambda = c_1 - c_2\delta^2$ implies that $\lambda + 1 < 0$. It follows from (26) that the squared coefficient of variation will be monotonically decreasing in k whenever $-(\lambda+1)$ is monotonically increasing (that is, when its derivative is positive) $\forall k \in (0, \frac{2nb-3a^2}{2nb+a^2})$.
To determine the sign of $\frac{d}{dk}\{c_2\delta^2 - (c_1+1)\}$, we focus on the sign of the numerator in $\frac{d}{dk}\frac{c_2 k\sigma^2(k\gamma^2\sigma^2 - 2c_1)}{1-2kc_2\sigma^2}$.

$$\big(2c_2k\gamma^2\sigma^4 - 2c_1c_2\sigma^2\big)\big(1 - 2kc_2\sigma^2\big) + \big(c_2k^2\gamma^2\sigma^4 - 2c_1c_2k\sigma^2\big)2c_2\sigma^2 > 0$$

may be simplified to

$$2c_2\sigma^2\big(k\gamma^2\sigma^2 - c_2k^2\gamma^2\sigma^4 - c_1\big) > 0.$$

This inequality holds whenever $k\gamma^2\sigma^2(1 - c_2k\sigma^2) > c_1$. Given that $c_1 < 0$, this condition holds $\forall k \in (0, \frac{2nb-3a^2}{2nb+a^2})$ because for k within this range, we know that $1 - 2c_2k\sigma^2 > 0$ implies that $1 - c_2k\sigma^2 > 0$. □

When selecting a value of k within its permissible range, we should consider that the larger the value of k, the closer the expected value of the prior is to the guess σ_0^2, but also the smaller the prior's coefficient of variation (that is the more informative is the distribution). Likewise, a larger a priori coefficient of variation yields an expected value farther to the left of σ_0^2.

The parameter k may be interpreted as a measure of the amount of a priori information available. If we are quite uncertain about our prior guess for σ^2 it makes sense to have not only a prior with a large coefficient of variation but also a conservatively small expected value. In fact, it can be easily shown through simulation that prior specification that put a sizeable part of the prior mass beyond σ^2 yields Bayes estimators with huge frequentist MSEs. Therefore, the relationship (through k) between a prior expected value and the coeffient of variation is consistent with the optimization of frequentist MSE of the Bayes estimators.

We also note that the relationship between k and the squared CV is non-linear. The speed of the squared CV reduction increases with k; therefore, choices of k close to its maximum imply very peaked prior distributions. On the other extreme, a small k leads to a prior that is practically the same as those considered in Section 4.1.

5 Simulation

In this section, we compare, using their frequentist MSEs, the Bayes estimators introduced in the previous sections with other estimators, both frequentist and Bayesian. We focus on the estimating the mean of the log-normal (i.e., $\theta_{1,0.5}$), a popular problem in the literature and for which several competing estimators exist. Specifically, we consider the unbiased estimator of Evans and Shaban (1976) and the estimator proposed by Shen et al. (2006) which the authors proved to be more efficient than the various alternatives that had previously been proposed in the literature. We also consider the Bayes estimator of Rukhin (1986). As a general benchmark, we also consider the MSE-optimal estimator of Zellner (1971) which is based on a known σ^2.

The simulation set-up is essentially the same as that introduced in Zhou and Gao (1997) and also used by Shen et al. (2006). We assume that $X \sim N(-\sigma^2, \sigma^2)$; then $\exp(X)$ is log-normally distributed, and our estimand $E\big[\exp(X)\big] = 1$. We consider 6 distinct values for σ^2 ($\sigma^2 = 0.1, 0.5, 1, 2, 5, 20$) and three sample sizes ($n = 11, 101, 400$). The results we present used $M = 100,000$ MC samples, and were obtained using `R`.

In table 1, we present the results obtained under the priors specified according to the suggestions of Section 4.1; specifically we set $\delta = 0.01$. In presenting the results, we denote the estimator of formula (3.1) in Rukhin (1986) by $\hat{\theta}^{Ru}$, with the prior parameters chosen using formulas (3.6) and (3.7) of the same paper. Let $\hat{\theta}^{ES}$ denote the unbiased estimator of Evans and Shaban (1976) and let $\hat{\theta}^{SBZ}$ denote the estimator from Shen et al. (2006). We also denote the optimal estimator assuming a known σ^2 that was discussed above by $\hat{\theta}^{Zel}$. Because the simulations only address estimating the mean, we omit the indexes related to the choice of a, b.

Despite the prior parameters having been chosen using a 'small arguments' approximation to, the Bessel-K function, table 1 shows that θ^{BQ} performs well regardless of the size of σ^2. Specifically θ^{BQ} is close to θ^{SBZ} for all values of σ^2, n. We emphasize this as θ^{SBZ} is the reference frequentist estimator in the recent literature. Both estimators are negatively biased and the bias rapidly increases with σ^2, especially when $n = 11$. In any case, note that θ^{Zel}, which is MSE-optimal and assumes known σ^2, is similar in this respect. θ^{BQ} is also more efficient than θ^{ES}. For small σ^2, θ^{BQ} is only moderately biased and shows a similar MSE, while for large population variances, the unbiasedeness of θ^{ES} is quite costly in terms of variance.

When comparing θ^{BQ} to θ^{BR}, we find that the two performs similarly for small σ^2, but θ^{BQ} becomes clearly superior to the predictor based on the relative quadratic loss function as σ^2 increases. θ^{Ru} exhibits an MSE close to that of θ^{BQ} for small σ^2, but its properties deteriorate dramatically for large population variances. We have already noted that the prior chosen for σ^2 in Rukhin (1986) is the limit of a GIG for $\delta \to 0$; nonetheless the proposed choice for γ is inconsistent with the existence of the first two moments of the posterior distribution for the mean of the log-normal. The results for $\hat{\theta}^{Ru}$ when $n = 400$ and $\sigma^2 \leq 2$ are missing because of numerical problems. These estimators involve the calculation of Bessel-K functions with very large orders (the order increases linearly with n) and arguments very close to 0, which results in huge values. These values lead to numerical instability and the generation of errors with the software we used. Because they are not essential to our purposes, we do not investigate the problem further.

In table 2, we report selected results for the estimators of the log-normal mean in the case where the priors for σ^2 incorporate prior guesses using the

Table 1: Comparison of alternative estimators: prior for σ^2 not incorporating guesses

n	Est. $/\sigma^2$	MSE						Bias					
		0.1	0.5	1	2	5	20	0.1	0.5	1	2	5	20
n=11	$\hat{\theta}^{BQ}$	0.0095	0.0514	0.109	0.228	0.531	0.982	-0.0075	-0.0470	-0.103	-0.225	-0.555	-0.990
	$\hat{\theta}^{BR}$	0.0096	0.0540	0.120	0.268	0.661	0.998	-0.0254	-0.1278	-0.246	-0.445	-0.793	-0.999
	$\hat{\theta}^{Ru}$	0.0095	0.0513	0.110	0.244	0.680	1.291	-0.0102	-0.0540	-0.107	-0.207	-0.448	-0.921
	$\hat{\theta}^{SBZ}$	0.0095	0.0505	0.105	0.217	0.544	0.998	-0.0108	-0.0621	-0.134	-0.288	-0.680	-0.999
	$\hat{\theta}^{ES}$	0.0098	0.0594	0.147	0.422	2.533	65.77	0.0010	0.0015	0.002	0.004	0.004	0.007
	$\hat{\theta}^{Zel}$	0.0093	0.0458	0.090	0.173	0.384	0.944	-0.0075	-0.0408	-0.082	-0.159	-0.354	-0.827
n=101	$\hat{\theta}^{BQ}$	0.0010	0.0060	0.014	0.036	0.133	0.729	-0.0014	-0.0067	-0.015	-0.039	-0.154	-0.812
	$\hat{\theta}^{BR}$	0.0010	0.0060	0.014	0.038	0.170	0.929	-0.0034	-0.0186	-0.042	-0.104	-0.341	-0.963
	$\hat{\theta}^{Ru}$	0.0011	0.0067	0.017	0.052	0.308	18.970	0.0029	0.0160	0.033	0.067	0.176	0.818
	$\hat{\theta}^{SBZ}$	0.0010	0.0060	0.014	0.036	0.134	0.805	-0.0023	-0.0111	-0.024	-0.055	-0.191	-0.888
	$\hat{\theta}^{ES}$	0.0010	0.0061	0.015	0.039	0.183	4.562	-0.0004	-0.0005	0.000	0.000	0.001	-0.014
	$\hat{\theta}^{Zel}$	0.0010	0.0049	0.010	0.020	0.486	0.182	-0.0015	-0.0061	-0.011	-0.022	-0.052	-0.185
n=401	$\hat{\theta}^{BQ}$	0.0003	0.0016	0.004	0.010	0.041	0.375	0.000	-0.002	-0.004	-0.010	-0.044	-0.415
	$\hat{\theta}^{BR}$	0.0003	0.0016	0.004	0.010	0.045	0.566	-0.001	-0.005	-0.011	-0.029	-0.118	-0.732
	$\hat{\theta}^{Ru}$					0.162	9.755					0.278	1.661
	$\hat{\theta}^{SBZ}$	0.0003	0.0016	0.004	0.010	0.041	0.376	-0.0006	-0.0029	-0.006	-0.015	-0.054	-0.447
	$\hat{\theta}^{ES}$	0.0003	0.0016	0.004	0.010	0.045	0.780	-0.0001	-0.0002	0.000	0.000	0.000	0.002
	$\hat{\theta}^{Zel}$	0.0003	0.0013	0.003	0.005	0.012	0.049	-0.0004	-0.0015	-0.003	-0.006	-0.013	-0.050

methodology described in Section 4.2. These priors are relevant only for large σ^2; therefore, we report the results only for $\sigma^2 = 1, 5, 20$. They involve a parameter k that controls both the closeness of the prior expectation to σ_0^2 (the prior guess) and the peakedeness of the prior. Let M denote the maximum admissible value for k according to (25); we consider the cases where $k = 0.5M$ and $k = M$. The purpose of this second set of comparisons involving $\hat{\theta}^{QB}$ is mainly to determine how much efficiency improves when a guess of the population variance is available and used in the prior specification and how sensitive these gains are to incorrect guesses for σ^2. For this purpose, we consider the following values for σ_0^2: $0.4\sigma^2, 0.8\sigma^2, \sigma^2, 1.2\sigma^2, 1.6\sigma^2, 2\sigma^2$. These values correspond to errors in guessing σ^2 that range from large and negative to large and positive. The results are reported only for $n = 11, 101$ because the impact of the prior specification is less interesting for $n = 400$.

With respect to Table 2, we note that, as expected, when the guess σ_0^2 is exactly equal to σ^2 there are large gains in effciency relative to the parallel results in Table 1. The gains are larger for $k = M$ than for $k = 0.5M$. To better appreciate the size of these gains, the MSEs and biases should be also be compared with those of $\hat{\theta}^{Zel}$ from Table 1, whose MSEs represent a lower bound.

When $k = 0.5M$, there is much less sensitivity to wrong guesses for σ^2 than when $k = M$. As anticipated in Section 4.2 this finding is explained by the prior becoming peaked and the prior coefficient of variation decreasing dramatically when k gets close to its upper bound. We observe this decreased sensitivity for both under- and over- statement of σ^2. Specifically, when $\sigma_0^2 = 0.4\sigma^2$, the prior implied by $k = 0.5M$ is more diffuse, which compensates for its expectation being farther to the right than in the case of $k = M$. For values of the guess close to σ^2, a more peaked distribution leads to estimators that behave better. However, as soon as the prior expected value exceeds the underlying true value (when $\sigma_0^2 = 1.2\sigma^2$, for example), we start to observe an increase in the MSE and a switch in the sign of the bias (from negative to positive) for the estimators associated with $k = M$. For larger positive errors in guessing σ^2, the properties of these estimators deteriorates fast and dramatically.

For $k = 0.5M$, we observe smoother behavior. The estimators with the best MSE are not those where $\sigma_0^2 = \sigma^2$. This results is due to $E(\sigma^2) = k\sigma_0^2$ which imply that a value of σ^2 moderately greater than σ^2 yields a prior with an expected value closer to σ^2. It is also unwise in this situation to intentionally overstate σ^2. From the case of $\sigma_0^2 = 2\sigma^2$, we may observe that the properties of $\hat{\theta}^{QB}$ when $k = 0.5M$ also deteriorate when there is gross positive error in guessing the population variance.

We conclude that when a reasonable prior guess for σ^2 is available, the efficiency of the $\hat{\theta}^{QB}$ Bayes estimators may be improved. Except for the case of k close to the upper limit of its admissible range, these improvements are substantial and can be obtained with a reasonable level of robustness with respect to guessing errors. Concerning the choice of k, we presented empirical results for $k = 0.5M$ that represent the best performance for the simulation setting

Table 2: Comparison of alternative estimators: prior incorporating guesses of σ^2

		MSE			*Bias*		
	n	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 20$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 20$
				$\sigma_0^2 = 0.4\sigma^2$			
$k = .5M$	11	0.101	0.521	0.988	-0.157	-0.638	-0.994
$k = M$	11	0.110	0.593	0.995	-0.227	-0.743	-0.997
$k = .5M$	101	0.014	0.137	0.792	-0.031	-0.226	-0.873
$k = M$	101	0.023	0.309	0.973	-0.117	-0.540	-0.986
				$\sigma_0^2 = 0.8\sigma^2$			
$k = .5M$	11	0.097	0.474	0.977	-0.121	-0.564	-0.987
$k = M$	11	0.091	0.432	0.952	-0.140	-0.564	-0.974
$k = .5M$	101	0.014	0.127	0.736	-0.021	-0.181	-0.830
$k = M$	101	0.013	0.131	0.772	-0.052	-0.310	-0.876
				$\sigma_0^2 = \sigma^2$			
$k = .5M$	11	0.096	0.453	0.963	-0.101	-0.515	-0.979
$k = M$	11	0.088	0.377	0.875	-0.092	-0.421	-0.913
$k = .5M$	101	0.014	0.123	0.680	-0.015	-0.149	-0.779
$k = M$	101	0.011	0.067	0.324	-0.013	-0.106	-0.498
	n			$\sigma_0^2 = 1.2\sigma^2$			
$k = .5M$	11	0.096	0.438	0.940	-0.080	-0.455	-0.964
$k = M$	11	0.091	0.406	0.986	-0.040	-0.227	-0.701
$k = .5M$	101	0.014	0.120	0.609	-0.009	-0.109	-0.693
$k = M$	101	0.013	0.131	2.934	0.029	0.189	1.238
				$\sigma_0^2 = 1.6\sigma^2$			
$k = .5M$	11	0.099	0.457	0.876	-0.036	-0.296	-0.879
$k = M$	11	0.117	1.316	80.534	0.074	0.401	2.648
$k = .5M$	101	0.014	0.132	0.683	0.005	-0.004	-0.281
$k = M$	101	0.030	1.845	3483.366	0.124	1.236	52.430
				$\sigma_0^2 = 2\sigma^2$			
$k = .5M$	11	0.108	0.627	1.666	0.012	-0.069	-0.539
$k = M$	11	0.181	6.355	13668.7	0.204	1.573	44.9
$k = .5M$	101	0.015	0.188	5.421	0.019	0.141	1.067
$k = M$	101	0.074	13.207	2684534	0.238	3.463	1468.1

considered here and also for others we do not report for the sake of brevity. A smaller ks tends to reproduce the results we have seen for the choices of prior parameters discussed in Section 4.1 while a larger k shares, although to a lesser degree, the problems illustrated for $k = M$.

6 Conclusions and future work

In this paper, we considered the popular log-normal model and the specific problems associated with estimating many of its parameters (including mean, median and mode). These problems are caused by the fact that $log-t$ and other distributions that can be met the analysis of the log-normal model have no finite moments. Our approach is Bayesian but parallel problems arise from a frequentist perspective. Specifically we wanted to continue using the popular quadratic loss function to summarize the posterior distribution. We found that a generalized inverse gamma prior for the population variance allows formally stating the conditions on the prior parameters that lead to posterior distributions with finite moments; moreover, the Bayes estimators of log-normal parameters associated to quadratic loss function have desirable frequentist properties.

Further developments of this research are possible in many directions. We are interested in applying generalized inverse gamma priors to the variance components in normal mixed models specified for the log of Poisson means that are commonly used (for example) in epidemiology. An extension of this methodology may also be applied to prediction problems in finite population modeling where normal models on the log-scale are popular. From a more theoretical point of view, we want to explore the relationship between our estimator and the other solutions proposed in the literature, such as the estimator of the log-normal mean proposed by Shen et al. (2006) which performs similarly to ours in simulations.

References

Abramowitz, A., and Stegun, I.A. (1968), Handbook of mathematical functions, ($5^t h$ edition), Dover, New York.

Barndorff-Nielsen, O.E. (1977), Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Statistical Society*, series A, 353, 401–419.

Bibby, B.M., and Sørensen M. (2003), Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes, in Rachev, S.T. (ed.) *Handbook of Heavy Tailed Distributions in Finance*, Elsevier Science B.V., 211–248.

Breymann, W., and Lüthi, D. (2010), `ghyp`: a package on Generalized Hyperbolic distributions, availalble at `cran.r-project.org/web/packages/ghyp/index.html`.

Eberlein, E., and von Hammerstein E.A. (2004), Hyperbolic processes in finance, in Dalang, R.C., Dozzi,M., Russo F. (eds.) (ed.) *Seminar on Stochastic Analysis, Random Fields and Applications IV, Progress in Probability*, Birkhäuser Verlag, 221–264.

Evans, I.G., and Shaban, S.A. (1976) New estimators (of smaller MSE) for parameters of a lognormal distribution, *Biometrische Zeitschrift*, 18, 453–466.

Gelman, A., (2006) Prior distributions for variance parameters in hierarchical models, *Bayesian Analysis*, 1, 515–533.

Gill, P.S., (2004) Samll-sample inference for the comparison of means of log-normal distributions, *Biometrics*, 60, 525–527.

Laforgia, A., and Natalini, P. (2010), Some inequalities for modified Bessel functions, *Journal of Inequalities and Applications*, Article ID 253035, 10 pages.

Limpert, E., Stahel, W.A., and Abbt M. (2001), Log-normal distributions across the sciences: keys and clues, *BioScience*, 51, 341–352.

Nguyen, T.T., Chen, J.T., Gupta, A.K., and Dinh, K.T. (2003) A proof of the conjecture on positive skeweness of Generalized Inverse Gaussian distributions, *Biometrika*, 90, 245–250.

Rukhin, A.L. (1986), Improved estimation in lognormal models, *Journal of the American Statistical Association*, 81, 1046–1049.

Shen, H., Brown, L.D., and Zhi H. (2006), Efficient estimation of log-normal means with application to pharmacokinetic data, *Statistics in Medicine*, 25, 3023–3038.

Zellner, A. (1971), Bayesian and non-Bayesian analysis of the log-normal distribution and log-normal regression, *Journal of the American Statistical Association*, 66, 327–330.

Zhou, X.H., and Gao, S. (1997) Confidence intervals of the log-normal mean, *Statistics in Medicine*, 16, 783–790.