

Article

Not peer-reviewed version

The Empirical Bayes Estimators of the Variance Parameter of the Normal Distribution with a Normal-Inverse-Gamma Prior under Stein's Loss Function

[Ying-Ying Zhang](#)*

Posted Date: 12 December 2025

doi: 10.20944/preprints202512.1148.v1

Keywords: empirical bayes estimators; hierarchical normal and normal-inverse-gamma model; maximum likelihood estimation (MLE) method; moment method; stein's loss function



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

The Empirical Bayes Estimators of the Variance Parameter of the Normal Distribution with a Normal-Inverse-Gamma Prior Under Stein's Loss Function

Ying-Ying Zhang ^{1,2}

¹ Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650500, China; yyzhang@ynu.edu.cn

² Department of Statistics and Data Science, School of Mathematics and Statistics, Yunnan University, Kunming 650500, China

Abstract

For the hierarchical normal and normal-inverse-gamma model, we derive the Bayesian estimator of the variance parameter in the normal distribution under Stein's loss function—a penalty function that treats gross overestimation and underestimation equally—and compute the associated Posterior Expected Stein's Loss (PESL). Additionally, we determine the Bayesian estimator of the same variance parameter under the squared error loss function, along with its corresponding PESL. We further develop empirical Bayes estimators for the variance parameter using a conjugate normal-inverse-gamma prior, employing both the method of moments and Maximum Likelihood Estimation (MLE). Through numerical simulations, we examine five key aspects: (1) the consistency of moment-based and MLE-based hyperparameter estimators; (2) the influence of κ_0 on quantities of interest as functions of the most recent observation; (3) two inequalities involving the Bayesian estimators and their respective PESL values; (4) the model's goodness-of-fit to simulated data; and (5) graphical representations of marginal densities under different hyperparameter settings. The simulation results demonstrate that MLEs outperform moment estimators in estimating hyperparameters, particularly with respect to consistency and model fit. Finally, we apply our methodology to real-world data on poverty levels—specifically, the percentage of individuals living below the poverty line—to validate and illustrate our theoretical findings.

Keywords: empirical bayes estimators; hierarchical normal and normal-inverse-gamma model; maximum likelihood estimation (MLE) method; moment method; stein's loss function

2020 MSC: 62C12; 62F10; 62F15

1. Introduction

The empirical Bayes method depends on a conjugate prior modeling, where the hyperparameters are estimated from the observations and the “estimated prior” is then used as a regular prior in the subsequent inference. See [1–3] and the references therein. The empirical Bayes method is introduced in [4–6]. From Bayesian perspective, it means that the sampling distribution is known, but the prior distribution is not. The marginal distribution is then used to restore the prior distribution from the observations. More literature on empirical Bayes method can be found, for instance, in [7–24].

The motivations of this paper are summarized as follows. Example 1.5.1 (p.20) of [25], Part I (pp.69–70) of [26], and [19] have considered the following hierarchical normal and normal-inverse-gamma model:

$$\begin{cases} X_i | (\mu, \theta) \stackrel{\text{iid}}{\sim} N(\mu, \theta), \quad i = 1, 2, \dots, n, \\ \mu | \theta \sim N(\mu_0, \theta/\kappa_0), \quad \theta \sim \widetilde{IG}(\nu_0/2, \nu_0\sigma_0^2/2), \end{cases} \quad (1)$$

where $-\infty < \mu_0 < \infty$, $\kappa_0 > 0$, $\nu_0 > 0$, and $\sigma_0^2 > 0$ are known hyperparameters, and iid means independent and identically distributed. The $N - \widetilde{IG}(\mu_0, \kappa_0, \nu_0, \sigma_0^2)$ distribution is a joint conjugate prior for (μ, θ) of the normal distribution $N(\mu, \theta)$, so that the posterior distribution of (μ, θ) is an $N - \widetilde{IG}(\mu_n, \kappa_n, \nu_n, \sigma_n^2)$ distribution with updated hyperparameters. However, in reality, the hyperparameters are unknown. [19] have estimated the hyperparameters of the model (1) by the moment method and the Maximum Likelihood Estimation (MLE) method. Moreover, they obtained the Bayes estimators of the mean and variance parameters of the model (1) under the squared error loss function. Finally, they obtained the empirical Bayes estimators of the mean and variance parameters of the model (1) under the squared error loss function by the moment method and the MLE method. However, in their empirical Bayes estimators, the sample $\mathbf{x}^{(n)} = (x_1, x_2, \dots, x_n)$ have been used twice. First, the sample $\mathbf{x}^{(n)}$ are utilized to estimate the hyperparameters μ_0 , κ_0 , ν_0 , and σ_0^2 . Second, the sample $\mathbf{x}^{(n)}$ are used to obtain the Bayes estimators. To avoid using the sample $\mathbf{x}^{(n)}$ twice, and to be compatible with the usual empirical Bayes analysis, we will use the following hierarchical normal and normal-inverse-gamma model in this paper:

$$\begin{cases} X_i | (\mu_i, \theta_i) \stackrel{\text{independent}}{\sim} N(\mu_i, \theta_i), i = 1, 2, \dots, n, n+1, \\ \mu_i | \theta_i \stackrel{\text{independent}}{\sim} N(\mu_0, \theta_i / \kappa_0), \theta_i \stackrel{\text{iid}}{\sim} \widetilde{IG}(\nu_0/2, \nu_0 \sigma_0^2/2), \end{cases} \quad (2)$$

where $-\infty < \mu_0 < \infty$, $\kappa_0 > 0$, $\nu_0 > 0$, and $\sigma_0^2 > 0$ are hyperparameters to be determined, $-\infty < \mu_{n+1} < \infty$ and $\theta_{n+1} > 0$ are the unknown parameters of interest, $N(\mu_i, \theta_i)$ is a normal distribution with an unknown mean μ_i and an unknown variance θ_i , the conditional conjugate prior distribution of μ_i given θ_i is $N(\mu_0, \theta_i / \kappa_0)$ which is a normal distribution with mean μ_0 and an unknown variance θ_i / κ_0 , the marginal conjugate prior distribution of θ_i is $\widetilde{IG}(\nu_0/2, \nu_0 \sigma_0^2/2)$ which is an inverse gamma distribution with shape parameter $\nu_0/2$ and scale parameter $\nu_0 \sigma_0^2/2$. Note that the joint conjugate prior $\pi(\mu, \theta) \sim N - \widetilde{IG}(\mu_0, \kappa_0, \nu_0, \sigma_0^2)$ is a normal-inverse-gamma distribution. As described in [7], the statistician observes data $(x_1, x_2, \dots, x_{n+1}) = \mathbf{x}^{(n+1)}$ and wishes to make an inference about μ_{n+1} and θ_{n+1} . Therefore, x_{n+1} provides direct information about the parameters μ_{n+1} and θ_{n+1} , while supplementary information $(x_1, x_2, \dots, x_n) = \mathbf{x}^{(n)}$ is also available. The connection between the prime data x_{n+1} and the supplementary information $\mathbf{x}^{(n)}$ is provided by the common distributions $N(\mu_i, \theta_i)$, $N(\mu_0, \theta_i / \kappa_0)$, and $\widetilde{IG}(\nu_0/2, \nu_0 \sigma_0^2/2)$. Moreover, since the variance parameter of the normal distribution is a positive restricted parameter, the squared error loss function is not appropriate. In contrast, we will choose Stein's loss function because it penalizes gross overestimation and gross underestimation equally, that is, an action a will incur an infinite loss when it tends to 0 or ∞ . Note that the squared error loss function does not have this property. For more literature on Stein's loss function, we refer readers to [20,27–30].

The key contributions of this work are outlined below. For the hierarchical normal and normal-inverse-gamma (N-NIG) framework (2), we begin by deriving the posterior distributions and marginal density, as stated in Theorem 1. Subsequently, we derive Bayes estimators for the variance parameter of the normal distribution under two loss criteria: Stein's loss and the squared error loss. We further compute the corresponding Posterior Expected Stein's Losses (PESLs) associated with these two estimators. It is shown that the two Bayes estimators and their respective PESLs satisfy specific inequalities, namely (15) and (16). Additionally, we derive moment-based estimators for the hyperparameters of the N-NIG model (2) and establish their consistency in Theorem 2. Similarly, Maximum Likelihood Estimators (MLEs) for the same hyperparameters are derived, and their consistency is proven in Theorem 3. Building on these results, we construct empirical Bayes estimators for the variance parameter under Stein's loss using both the method of moments and MLE, as presented in Theorem 4. The theoretical findings are supported by numerical experiments and an application to real-world data.

Comparing models (1) and (2) carefully, we find that the samples $\mathbf{x}_1^{(n)} = (x_{11}, x_{12}, \dots, x_{1n})$ and $\mathbf{x}_2^{(n)} = (x_{21}, x_{22}, \dots, x_{2n})$ generated from the two models are iid from different distributions. On the one hand, the sample $\mathbf{x}_1^{(n)}$ generated from (1) are iid from $N(\mu, \theta)$. Although the marginal densities

of $x_{11}, x_{12}, \dots, x_{1n}$ are $t_{\nu_0}(\mu_0, \sigma_0^2(\kappa_0 + 1)/\kappa_0)$, and thus $x_{11}, x_{12}, \dots, x_{1n}$ can be thought to be from the t distribution. However, $x_{11}, x_{12}, \dots, x_{1n}$ are not iid from the t distribution. In other words, $x_{11}, x_{12}, \dots, x_{1n}$ are dependent from the t distribution. On the other hand, the sample $x_2^{(n)}$ generated from (2) are iid from the t distribution. That is, $x_{21}, x_{22}, \dots, x_{2n}$ are independent and identically distributed from the t distribution. The sample $x_1^{(n)}$ can be used to estimate the parameters μ and θ from $N(\mu, \theta)$, while the sample $x_2^{(n)}$ can be used to estimate the hyperparameters μ_0, ν_0 , and u_0 from $t_{\nu_0}(\mu_0, u_0)$, where $u_0 = \sigma_0^2(\kappa_0 + 1)/\kappa_0$.

The structure of this paper is outlined as follows. Section 2 begins with the derivation of posterior distributions and the marginal likelihood for the hierarchical normal and normal-inverse-gamma model. Subsequently, we derive the corresponding Bayes estimators and PESLs, which are shown to fulfill two key inequalities: (15) and (16). Additionally, Theorem 4 presents the empirical Bayes estimators for the model's variance component under Stein's loss function, obtained via both the method of moments and MLE. In Section 3, we conduct numerical experiments from five different perspectives to evaluate performance. Section 4 illustrates the methodology using real-world data, specifically poverty rate data reflecting the proportion of individuals living below the poverty line. Finally, Section 5 offers concluding remarks and a discussion of the findings.

2. Theoretical Results

Suppose that $X \sim \widetilde{G}(\alpha, \beta)$ and $Y = 1/X \sim \widetilde{IG}(\alpha, \beta)$. The probability density functions (pdfs) of X and Y are respectively given by

$$f_X(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-x\beta), \quad x > 0, \alpha, \beta > 0,$$

$$f_Y(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{y}\right)^{\alpha+1} \exp\left(-\frac{\beta}{y}\right), \quad y > 0, \alpha, \beta > 0.$$

It is easy to calculate $EX = \alpha/\beta$ and $EY = \beta/(\alpha - 1)$, for $\alpha > 1$ and $\beta > 0$.

2.1. The Bayes Estimators and the PESLs

For the hierarchical normal and normal-inverse-gamma model (2), we have the following theorem which calculates the posterior densities $\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1})$, $\pi(\theta_{n+1}|x_{n+1})$, $\pi(\mu_{n+1}|x_{n+1})$, $\pi(\mu_{n+1}|\theta_{n+1}, x_{n+1})$, and the marginal density $m(x_{n+1}|\mu_0, \kappa_0, \nu_0, \sigma_0^2)$. The proof of the theorem can be found in the appendix.

Theorem 1. For the hierarchical normal and normal-inverse-gamma model (2), the joint posterior density of $(\mu_{n+1}, \theta_{n+1})$ is

$$\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1}) \sim N - \widetilde{IG}(\mu_1, \kappa_1, \nu_1, \sigma_1^2),$$

the marginal posterior density of θ_{n+1} is

$$\pi(\theta_{n+1}|x_{n+1}) \sim \widetilde{IG}(\alpha^*, \beta^*),$$

the marginal posterior density of μ_{n+1} is

$$\pi(\mu_{n+1}|x_{n+1}) \sim t_{\nu_1} \left(\mu_1, \frac{\sigma_1^2}{\kappa_1} \right),$$

the conditional posterior density of μ_{n+1} is

$$\pi(\mu_{n+1}|\theta_{n+1}, x_{n+1}) \sim N(\mu_1, \theta_{n+1}/\kappa_1),$$

where

$$\begin{cases} \mu_1 = (x_{n+1} + \kappa_0 \mu_0) / (\kappa_0 + 1), \\ \kappa_1 = \kappa_0 + 1, \\ \nu_1 = \nu_0 + 1, \\ \nu_1 \sigma_1^2 = \nu_0 \sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1} (x_{n+1} - \mu_0)^2, \end{cases} \quad (3)$$

$$\alpha^* = \frac{\nu_1}{2} = \frac{\nu_0 + 1}{2}, \quad (4)$$

and

$$\beta^* = \frac{\nu_1 \sigma_1^2}{2} = \frac{1}{2} \left[\nu_0 \sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1} (x_{n+1} - \mu_0)^2 \right]. \quad (5)$$

Moreover, the marginal density of x_{n+1} is given by

$$m(x_{n+1} | \mu_0, \kappa_0, \nu_0, \sigma_0^2) \sim t_{\nu_0} \left(\mu_0, \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} \right),$$

with probability density function (pdf) given by

$$\begin{aligned} m(x_{n+1} | \mu_0, \kappa_0, \nu_0, \sigma_0^2) \\ = \frac{\Gamma\left(\frac{\nu_0 + 1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right) \sqrt{\pi \nu_0 \sigma_0^2 (\kappa_0 + 1) / \kappa_0}} \left[1 + \frac{1}{\nu_0} \frac{(x_{n+1} - \mu_0)^2}{\sigma_0^2 (\kappa_0 + 1) / \kappa_0} \right]^{-\frac{\nu_0 + 1}{2}}, \end{aligned} \quad (6)$$

for $-\infty < x_{n+1} < \infty$, $-\infty < \mu_0 < \infty$, $\kappa_0 > 0$, $\nu_0 > 0$, and $\sigma_0^2 > 0$.

In the following, we will calculate the Bayes estimator of θ_{n+1} under Stein's loss function $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$, the Bayes estimator of θ_{n+1} under the usual squared error loss function $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$, and the PESLs at $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$ ($PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$).

Similar to [28], the two Bayes estimators and the two PESLs are respectively given by

$$\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) = \frac{1}{E\left(\frac{1}{\theta_{n+1}} | x_{n+1}\right)}, \quad (7)$$

$$\delta_2^{\pi, \theta_{n+1}}(x_{n+1}) = E(\theta_{n+1} | x_{n+1}), \quad (8)$$

$$PESL_s^{\pi, \theta_{n+1}}(x_{n+1}) = \log E\left(\frac{1}{\theta_{n+1}} | x_{n+1}\right) + E(\log \theta_{n+1} | x_{n+1}), \quad (9)$$

$$\begin{aligned} PESL_2^{\pi, \theta_{n+1}}(x_{n+1}) &= E(\theta_{n+1} | x_{n+1}) E\left(\frac{1}{\theta_{n+1}} | x_{n+1}\right) - 1 \\ &\quad - \log E(\theta_{n+1} | x_{n+1}) + E(\log \theta_{n+1} | x_{n+1}). \end{aligned} \quad (10)$$

To calculate the two Bayes estimators and the two PESLs, it remains to calculate

$$E_1 = E(\theta_{n+1} | x_{n+1}), \quad E_2 = E\left(\frac{1}{\theta_{n+1}} | x_{n+1}\right), \quad \text{and} \quad E_3 = E(\log \theta_{n+1} | x_{n+1}).$$

Similar to [28], it can be shown that

$$\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) \leq \delta_2^{\pi, \theta_{n+1}}(x_{n+1}) \quad (11)$$

by exploiting Jensen's inequality. Moreover,

$$PESL_s^{\pi, \theta_{n+1}}(x_{n+1}) \leq PESL_2^{\pi, \theta_{n+1}}(x_{n+1}), \quad (12)$$

which is a direct consequence of the general methodology for finding a Bayes estimator. According to construction, $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ minimizes the Posterior Expected Stein's Loss (PESL).

Now let us analytically calculate the Bayes estimators $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$, and the PESLs $PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$ under the hierarchical normal and normal-inverse-gamma model (2) from (7), (8), (9), and (10). The three expectations are calculated as

$$\begin{aligned} E(\theta_{n+1}|x_{n+1}) &= \frac{\beta^*}{\alpha^* - 1}, \\ E\left(\frac{1}{\theta_{n+1}}|x_{n+1}\right) &= \frac{\alpha^*}{\beta^*}, \\ E(\log \theta_{n+1}|x_{n+1}) &= \log \beta^* - \psi(\alpha^*), \end{aligned}$$

for $\alpha^* > 1$, where α^* and β^* are given by (4) and (5). From (7), the Bayes estimator of θ_{n+1} under Stein's loss function is given by

$$\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) = \frac{1}{E\left(\frac{1}{\theta_{n+1}}|x_{n+1}\right)} = \frac{\beta^*}{\alpha^*}. \quad (13)$$

From (8), the Bayes estimator of θ_{n+1} under the usual squared error loss function is given by

$$\delta_2^{\pi, \theta_{n+1}}(x_{n+1}) = E(\theta_{n+1}|x_{n+1}) = \frac{\beta^*}{\alpha^* - 1}, \quad (14)$$

for $\alpha^* > 1$. It is easy to show that

$$\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) = \frac{\beta^*}{\alpha^*} < \frac{\beta^*}{\alpha^* - 1} = \delta_2^{\pi, \theta_{n+1}}(x_{n+1}), \quad (15)$$

which exemplifies the theoretical study of (11). Furthermore, from (9) and (10), the PESL at $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$ are respectively given by

$$PESL_s^{\pi, \theta_{n+1}}(x_{n+1}) = \log \alpha^* - \psi(\alpha^*)$$

and

$$PESL_2^{\pi, \theta_{n+1}}(x_{n+1}) = \frac{1}{\alpha^* - 1} + \log(\alpha^* - 1) - \psi(\alpha^*),$$

where

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d}{dx} \log \Gamma(x) = \text{digamma}(x)$$

is the digamma function. It can be shown that

$$\begin{aligned} PESL_s^{\pi, \theta_{n+1}}(x_{n+1}) &= \log \alpha^* - \psi(\alpha^*) \\ &\leq \frac{1}{\alpha^* - 1} + \log(\alpha^* - 1) - \psi(\alpha^*) \\ &= PESL_2^{\pi, \theta_{n+1}}(x_{n+1}), \end{aligned} \quad (16)$$

which exemplifies the theoretical study of (12). It is worth noting that the PESLs $PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$ depend only on α^* , but not on β^* . Therefore, the PESLs depend only on ν_0 , but not on $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} .

We will exemplify the two inequalities (15) and (16) in the simulations section and the real data section. Moreover, we will exemplify that the PESLs depend only on ν_0 , but not on $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} .

2.2. The Empirical Bayes Estimators of θ_{n+1}

The hyperparameters of model (2) are $-\infty < \mu_0 < \infty, \kappa_0 > 0, \nu_0 > 0$, and $\sigma_0^2 > 0$. However, we can not directly obtain the estimators of the four hyperparameters of model (2) by the moment method. Let

$$u_0 = \frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0}. \quad (17)$$

Since κ_0 and σ_0 appear together in u_0 , we can not directly obtain the estimators of κ_0 and σ_0 by the moment method. In other words, κ_0 and σ_0 are unidentifiable. In the empirical Bayesian statistical literature, common approaches to addressing the issue of unidentifiability of hyperparameters include the following two. One is to estimate the prior distribution through non-parametric or semi-parametric methods, avoiding strong assumptions about the functional form of the prior distribution, thereby circumventing the problem of unidentifiability of hyperparameters ([31,32]). The other is to use auxiliary data, model structure constraints, or specific assumptions (such as sparsity, spatial correlation) to provide additional information, making the unidentifiable hyperparameters identifiable ([33–35]). We adopt the second approach to make hyperparameters (κ_0 and σ_0^2) identifiable. More specifically, when κ_0 is fixed to be a known constant (our recommendation is $\kappa_0 = 1$, which will be made clear later in this subsection), then κ_0 and σ_0^2 are identifiable. Otherwise, κ_0 and σ_0^2 are unidentifiable.

However, we can obtain the estimator of u_0 by the moment method. In the following, we are interested in the hyperparameters $-\infty < \mu_0 < \infty, \nu_0 > 0$, and $u_0 > 0$. Using hyperparameters μ_0, ν_0 , and u_0 , the marginal density (6) changes to

$$\begin{aligned} & m(x_{n+1} | \mu_0, \nu_0, u_0) \\ &= \frac{\Gamma\left(\frac{\nu_0+1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right) \sqrt{\pi \nu_0 u_0}} \left[1 + \frac{1}{\nu_0} \frac{(x_{n+1} - \mu_0)^2}{u_0} \right]^{-\frac{\nu_0+1}{2}} \\ &\sim t_{\nu_0}(\mu_0, u_0). \end{aligned} \quad (18)$$

The moment-based estimators for the hyperparameters μ_0, ν_0 , and u_0 in model (2), denoted as $\mu_{01}(n), \nu_{01}(n)$, and $u_{01}(n)$, along with their consistency properties, are presented in the subsequent theorem. The proof of this theorem is provided in the appendix.

Theorem 2. *The estimators of the hyperparameters μ_0, ν_0 , and u_0 of model (2) by the moment method are*

$$\begin{aligned} \mu_{01}(n) &= A_1, \\ \nu_{01}(n) &= \frac{-\frac{14}{3}A_1^4 + 4A_1^2A_2 + 2A_2^2 - \frac{4}{3}A_4}{-\frac{2}{3}A_1^4 + A_2^2 - \frac{1}{3}A_4}, \\ u_{01}(n) &= \frac{-\frac{1}{3}(A_2 - A_1^2)(5A_1^4 - 6A_1^2A_2 + A_4)}{-\frac{7}{3}A_1^4 + 2A_1^2A_2 + A_2^2 - \frac{2}{3}A_4}, \end{aligned}$$

where

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, 4,$$

is the sample k th moment of X . Moreover, the moment estimators are consistent estimators of the hyperparameters.

We remark that the moment estimators $\mu_{01}(n)$, $\nu_{01}(n)$, and $u_{01}(n)$ in Theorem 2 are the same as those in Theorem 2 in [19]. The reason for the same moment estimators is that for the two hierarchical normal and normal-inverse-gamma models (1) and (2), the marginal distributions are the same, and the population moments of X are the same. Moreover, in Theorem 2 of this paper, we have shown that the moment estimators are consistent estimators of the hyperparameters, and this result has not been derived in [19].

The MLE method is used to derive estimators for the hyperparameters μ_0 , ν_0 , and u_0 of the model (2), denoted as $\mu_{02}(n)$, $\nu_{02}(n)$, and $u_{02}(n)$, respectively. The consistency properties of these estimators are presented in the subsequent theorem, with the detailed proof provided in the appendix.

Theorem 3. *The estimators of the hyperparameters μ_0 , ν_0 , and u_0 of the model (2) by the MLE method $\mu_{02}(n)$, $\nu_{02}(n)$, and $u_{02}(n)$ are the solutions to the following equations:*

$$f_1(\mu_0, \nu_0, u_0) = \frac{1}{n} \sum_{i=1}^n \frac{(\nu_0 + 1)(x_i - \mu_0)}{\nu_0 u_0 + (x_i - \mu_0)^2} = 0, \quad (19)$$

$$f_2(\mu_0, \nu_0, u_0) = \frac{1}{2} \psi\left(\frac{\nu_0 + 1}{2}\right) - \frac{1}{2} \psi\left(\frac{\nu_0}{2}\right) - \frac{1}{2\nu_0} - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \log \left[1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0} \right] - \frac{(\nu_0 + 1)(x_i - \mu_0)^2}{[\nu_0 u_0 + (x_i - \mu_0)^2] \nu_0} \right\} = 0, \quad (20)$$

$$f_3(\mu_0, \nu_0, u_0) = -\frac{1}{2u_0} + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(\nu_0 + 1)(x_i - \mu_0)^2}{[\nu_0 u_0 + (x_i - \mu_0)^2] u_0} = 0. \quad (21)$$

Moreover, the MLEs are consistent estimators of the hyperparameters.

The MLEs of the hyperparameters μ_0 , ν_0 , and u_0 cannot be derived analytically by solving equations (19), (20), and (21). As a result, numerical methods must be employed. Newton's method can be applied to solve these equations and thereby obtain the MLEs of the hyperparameters. It should be emphasized that the accuracy of the resulting MLEs is highly sensitive to the choice of initial values, and moment-based estimators have generally been found to serve as effective starting points.

Finally, the empirical Bayes estimates of the model (2)'s variance parameter under Stein's loss function, derived using the method of moments and the MLE, are presented in the subsequent theorem.

Theorem 4. *The empirical Bayes estimator of the variance parameter of the model (2) under Stein's loss function by the moment method is given by (13) with the hyperparameters estimated by $(\mu_{01}(n), \nu_{01}(n), u_{01}(n))$ in Theorem 2. Alternatively, the empirical Bayes estimator of the variance parameter of the model (2) under Stein's loss function by the MLE method is given by (13) with the hyperparameters estimated by $(\mu_{02}(n), \nu_{02}(n), u_{02}(n))$ numerically determined in Theorem 3.*

Now let us discuss the selection of κ_0 . We recommend choosing $\kappa_0 = 1$, and the reason is given as follows. From (13), (4), and (5), we have

$$\begin{aligned}\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) &= \frac{\beta^*}{\alpha^*} \\ &= \frac{\frac{1}{2} \left[v_0 \sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1} (x_{n+1} - \mu_0)^2 \right]}{\frac{v_0 + 1}{2}} \\ &= \frac{1}{v_0 + 1} \left[v_0 \sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1} (x_{n+1} - \mu_0)^2 \right] \\ &= \frac{\kappa_0}{\kappa_0 + 1} \frac{1}{v_0 + 1} \left[v_0 \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} + (x_{n+1} - \mu_0)^2 \right] \\ &= \frac{\kappa_0}{\kappa_0 + 1} \frac{1}{v_0 + 1} \left[v_0 u_0 + (x_{n+1} - \mu_0)^2 \right].\end{aligned}\quad (22)$$

It is easy to show that the factor

$$\frac{\kappa_0}{\kappa_0 + 1} \in (0, 1),$$

for $\kappa_0 > 0$. Because we have little information about κ_0 , we will choose

$$\frac{\kappa_0}{\kappa_0 + 1} = \frac{1}{2},$$

which is in the middle of the above range. Hence,

$$\kappa_0 = 1. \quad (23)$$

Therefore, (22) reduces to

$$\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) = \frac{1}{2} \frac{1}{v_0 + 1} \left[v_0 u_0 + (x_{n+1} - \mu_0)^2 \right],$$

which can then be estimated once the hyperparameters (μ_0, v_0, u_0) are estimated. From (17) and (23), we have

$$\sigma_0^2 = \frac{\kappa_0}{\kappa_0 + 1} u_0 = \frac{1}{2} u_0,$$

which can then be estimated once the hyperparameter u_0 is estimated.

Another reason to choose $\kappa_0 = 1$ is given below. Since $-\infty < \mu_{n+1} < \infty$, the squared error loss function is appropriate. The Bayes estimator of μ_{n+1} under the squared error loss function is given by

$$\delta_2^{\pi, \mu_{n+1}}(x_{n+1}) = E(\mu_{n+1} | x_{n+1}) = \mu_1 = \frac{x_{n+1} + \kappa_0 \mu_0}{\kappa_0 + 1}. \quad (24)$$

In (24), κ_0 represents a strength of belief on μ_0 . If one harbors no belief on μ_0 , then $\kappa_0 = 0$, and thus $\delta_2^{\pi, \mu_{n+1}}(x_{n+1}) = x_{n+1}$, which depends only on the datum x_{n+1} . In contrast, if one harbors complete belief on μ_0 , then $\kappa_0 = \infty$, and thus $\delta_2^{\pi, \mu_{n+1}}(x_{n+1}) = \mu_0$, which depends only on μ_0 . However, if one believes that x_{n+1} and μ_0 are equally important, then $\kappa_0 = 1$ is a reasonable choice, and thus $\delta_2^{\pi, \mu_{n+1}}(x_{n+1}) = (x_{n+1} + \mu_0)/2$, which is a balanced combination of x_{n+1} and μ_0 .

We remark that κ_0 and σ_0^2 affect β^* , $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$, $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$, and $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$, and they do not affect α^* , $PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$, and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$.

3. Simulations

This section presents numerical simulations for the hierarchical normal and normal-inverse-gamma model (2). We examine five key aspects: the consistency of moment estimators and MLEs, the influence of κ_0 on quantities of interest as functions of x_{n+1} , the validity of two inequalities concerning

Bayes estimators and PESLs, the model's goodness-of-fit to the generated data, and the marginal density distributions of the model parameters.

The simulated data are generated according to the hierarchical normal and normal-inverse-gamma model (2) with the hyperparameters specified by $\mu_0 = 3$, $\kappa_0 = 1$, $\nu_0 = 6$, and $\sigma_0^2 = 1$. The reason why we choose these values is that $-\infty < \mu_0 < \infty$, $\kappa_0 > 0$, $\nu_0 > 0$, and $\sigma_0^2 > 0$. Moreover, $\nu_0 > 4$ is required in moment estimations of the hyperparameters. Additional numerical values for the hyperparameters can likewise be defined.

3.1. Consistencies of the Moment Estimators and the MLEs

Currently, both the moment estimators $(\mu_{01}(n), \nu_{01}(n), u_{01}(n))$ and the MLEs $(\mu_{02}(n), \nu_{02}(n), u_{02}(n))$ serve as consistent estimates for the hyperparameters (μ_0, ν_0, u_0) of the hierarchical normal and normal-inverse-gamma model (2). The purpose of this subsection is to emphasize that, as established in Theorems 2 and 3, we have rigorously proven the consistency of both the moment-based and MLE-based estimators with respect to the true hyperparameters. It should be noted that all analyses here are based solely on the observed data $\mathbf{x}^{(n)} = (x_1, x_2, \dots, x_n)$.

First, we will numerically exemplify that the sample $\mathbf{x}_1^{(n)} = (x_{11}, x_{12}, \dots, x_{1n})$ generated from the model (1) can not be used to estimate the hyperparameters (μ_0, ν_0, u_0) , while the sample $\mathbf{x}_2^{(n)} = (x_{21}, x_{22}, \dots, x_{2n})$ generated from the model (2) can be used to estimate the hyperparameters (μ_0, ν_0, u_0) , where $u_0 = \sigma_0^2(\kappa_0 + 1)/\kappa_0$. Moreover, we will exemplify that the moment estimators and the MLEs of u_0 can correctly estimate the true hyperparameter u_0 regardless of the κ_0 and σ_0^2 values.

The histograms of the samples and their density curves are plotted in Figure 1. From the figure, we observe the following facts.

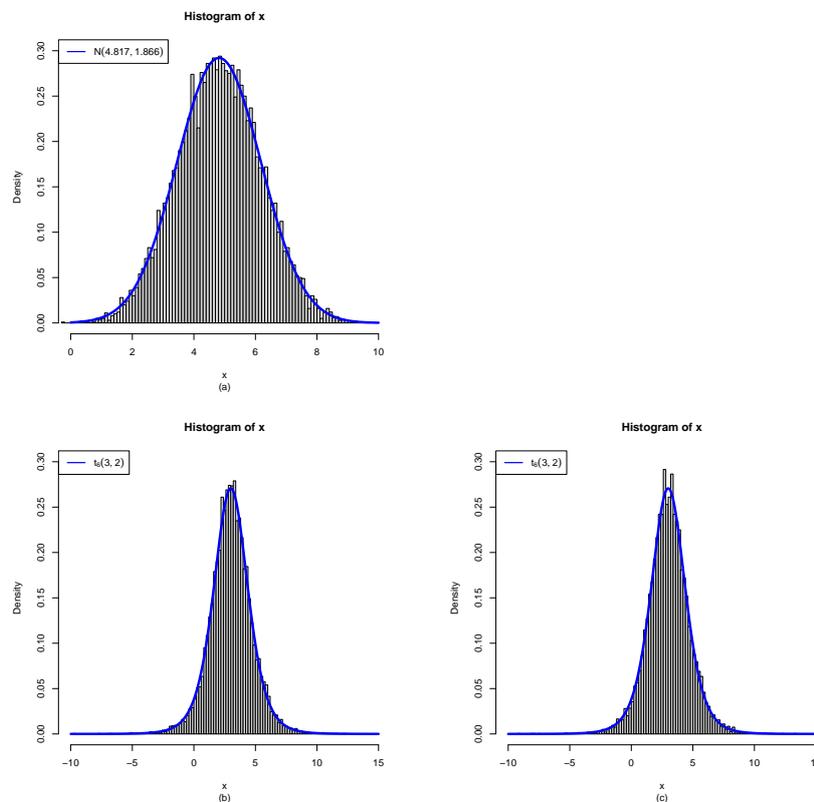


Figure 1. The histograms of the samples and their density curves. (a) $\mathbf{x}_1^{(n)}$ generated from (1) with $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 1, 6, 1)$ and $u_0 = 2$. (b) $\mathbf{x}_2^{(n)}$ generated from (2) with $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 1, 6, 1)$ and $u_0 = 2$. (c) $\mathbf{x}_3^{(n)}$ generated from (2) with $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 2, 6, 4/3)$ and $u_0 = 2$.

1. Plot (a): the sample $\mathbf{x}_1^{(n)}$ generated from (1) are iid from $N(\mu, \theta)$ with $\mu = 4.817$ and $\theta = 1.866$.

2. Plots (b) and (c): the sample $x_2^{(n)}$ are generated from (2) with $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 1, 6, 1)$ and $u_0 = 2$. The sample $x_3^{(n)}$ are generated from (2) with $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 2, 6, 4/3)$ and $u_0 = 2$. Although $(\kappa_0, \sigma_0^2) = (1, 1)$ and $(2, 4/3)$ are different in the two plots, $u_0 = 2$ is the same. Therefore, the two samples $x_2^{(n)}$ and $x_3^{(n)}$ are from the same marginal distribution $t_{\nu_0}(\mu_0, u_0)$ with $(\mu_0, \nu_0, u_0) = (3, 6, 2)$.

The moment estimators and the MLEs of the hyperparameters $(\mu_0, \nu_0, u_0) = (3, 6, 2)$ for the samples $x_1^{(n)}$, $x_2^{(n)}$, and $x_3^{(n)}$ are summarized in Table 1. From the table, we observe the following facts.

Table 1. The moment estimators and the MLEs of the hyperparameters $(\mu_0, \nu_0, u_0) = (3, 6, 2)$ for the samples $x_1^{(n)}$, $x_2^{(n)}$, and $x_3^{(n)}$.

	Moment estimators	MLEs
$x_1^{(n)}$	(4.826, -159.725, 1.872)	(4.826, -159.725, 1.872)
$x_2^{(n)}$	(3.025, 7.306, 2.127)	(3.020, 6.011, 1.967)
$x_3^{(n)}$	(3.021, 7.367, 2.138)	(3.018, 6.179, 1.990)

1. The moment estimators of the hyperparameters (μ_0, ν_0, u_0) for sample $x_1^{(n)}$ are far away from the true hyperparameters $(3, 6, 2)$, and thus the samples generated from the model (1) can not be used to estimate the hyperparameters (μ_0, ν_0, u_0) .
2. For $x_1^{(n)}$, since the moment estimator of ν_0 is -159.725 which is negative, the MLE method fails to iterate, and thus the MLEs of the hyperparameters (μ_0, ν_0, u_0) are equal to the moment estimators.
3. For $x_2^{(n)}$ and $x_3^{(n)}$, the moment estimators and the MLEs of the hyperparameters (μ_0, ν_0, u_0) are close to the true hyperparameters $(3, 6, 2)$, and thus the samples generated from the model (2) can be used to estimate the hyperparameters (μ_0, ν_0, u_0) .
4. The sample $x_2^{(n)}$ is generated from the model (2) with $(\kappa_0, \sigma_0^2) = (1, 1)$, while the sample $x_3^{(n)}$ is generated from the model (2) with $(\kappa_0, \sigma_0^2) = (2, 4/3)$. Although $(\kappa_0, \sigma_0^2) = (1, 1)$ and $(2, 4/3)$ are different for the two samples $x_2^{(n)}$ and $x_3^{(n)}$, $u_0 = 2$ is the same. We find that both the moment method and the MLE method correctly estimate the true hyperparameter $u_0 = 2$ for the two samples $x_2^{(n)}$ and $x_3^{(n)}$.
5. For $x_2^{(n)}$ and $x_3^{(n)}$, the MLEs are closer to the true hyperparameters $(\mu_0, \nu_0, u_0) = (3, 6, 2)$ than the moment estimators for this simulation.

Following the approach in [21], let ζ represent one of the hyperparameters μ_0 , ν_0 , or u_0 . Consistency is defined as the property that

$$\zeta_i(n) \xrightarrow{p} \zeta \quad \text{as } n \rightarrow \infty,$$

for $i = 1, 2$, where $i = 1$ corresponds to the moment estimator and $i = 2$ to the MLE, with \xrightarrow{p} denoting convergence in probability and n being the sample size. Equivalently, consistency can be expressed as

$$\lim_{n \rightarrow \infty} P(|\zeta_i(n) - \zeta| \geq \varepsilon) = 0,$$

for any $\varepsilon > 0$ and for all $\zeta \in \Xi$. These probabilities are estimated using empirical frequencies from simulations:

$$P(|\zeta_i(n) - \zeta| \geq \varepsilon) \approx \frac{1}{M} \sum_{m=1}^M I(|\zeta_i^{(m)}(n) - \zeta| \geq \varepsilon) \triangleq F_{\zeta_i}^{\varepsilon, M}(n),$$

where $I(A)$ is the indicator function taking value 1 if event A occurs and 0 otherwise, and M denotes the number of simulation replications. Hence, a decreasing trend in the simulated frequencies $F_{\zeta_i}^{\varepsilon, M}(n)$ toward zero as n increases provides empirical evidence of estimator consistency, for both estimation methods ($i = 1, 2$) and across all hyperparameters $\zeta \in \{\mu_0, \nu_0, u_0\}$.

We now demonstrate that the moment estimators $(\mu_{01}(n), \nu_{01}(n), u_{01}(n))$ and the MLEs $(\mu_{02}(n), \nu_{02}(n), u_{02}(n))$ are consistent estimators of the hyperparameters (μ_0, ν_0, u_0) . Table 2 presents the frequency distributions of these estimators for both moment-based and MLE for varying sample sizes n , under settings where $\varepsilon = 1, 0.5$, and 0.1 . The simulated data are generated from the hierarchical normal and normal-inverse-gamma model (2), with hyperparameter values set to $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 1, 6, 1)$ and $u_0 = 2$. While other configurations of these hyperparameters could also be considered, the current choice serves to illustrate the behavior effectively. Analysis of the table reveals several key observations.

Table 2. The frequencies of the moment estimators and the MLEs of the hyperparameters as n varies for $\varepsilon = 1, 0.5$, and 0.1 .

	n	Moment estimators			MLEs		
		$F_{\mu_{01}}^{\varepsilon, M}(n)$	$F_{\nu_{01}}^{\varepsilon, M}(n)$	$F_{u_{01}}^{\varepsilon, M}(n)$	$F_{\mu_{02}}^{\varepsilon, M}(n)$	$F_{\nu_{02}}^{\varepsilon, M}(n)$	$F_{u_{02}}^{\varepsilon, M}(n)$
$\varepsilon = 1$	1e4	0.00	0.24	0.00	0.00	0.03	0.00
	2e4	0.00	0.23	0.00	0.00	0.00	0.00
	4e4	0.00	0.06	0.00	0.00	0.00	0.00
	8e4	0.00	0.01	0.00	0.00	0.00	0.00
	16e4	0.00	0.00	0.00	0.00	0.00	0.00
$\varepsilon = 0.5$	1e4	0.00	0.48	0.00	0.00	0.16	0.00
	2e4	0.00	0.40	0.00	0.00	0.04	0.00
	4e4	0.00	0.33	0.00	0.00	0.00	0.00
	8e4	0.00	0.19	0.00	0.00	0.00	0.00
	16e4	0.00	0.10	0.00	0.00	0.00	0.00
$\varepsilon = 0.1$	1e4	0.00	0.81	0.36	0.00	0.84	0.03
	2e4	0.00	0.87	0.34	0.00	0.62	0.00
	4e4	0.00	0.82	0.21	0.00	0.52	0.00
	8e4	0.00	0.81	0.10	0.00	0.43	0.00
	16e4	0.00	0.70	0.04	0.00	0.27	0.00

1. With $\varepsilon = 1, 0.5$, or 0.1 , the frequencies of the estimators approach zero as n tends to infinity, indicating that both the moment estimators and the MLEs are consistent for estimating the hyperparameters. In the case of $\varepsilon = 0.1$, the frequencies of $\nu_{01}(n)$ and $\nu_{02}(n)$ remain relatively high (at least 0.2 across all scenarios). Nevertheless, a decreasing trend toward zero is evident as n grows large, suggesting eventual convergence.
2. By comparing the frequencies corresponding to $\varepsilon = 1, 0.5$, and 0.1 , we find that as ε decreases, the frequencies generally increase. This occurs because the constraints

$$|\tilde{\zeta}_i^{(m)}(n) - \zeta| \geq \varepsilon, \quad \text{for } \zeta = \mu_0, \nu_0, \text{ and } u_0, i = 1, 2,$$

become easier to satisfy when ε is smaller.

3. When comparing the moment estimators with the MLEs of the hyperparameters μ_0, ν_0 , and u_0 , it is observed that for large sample sizes n , the MLEs exhibit smaller values than the moment estimators. This indicates that, in terms of consistency, the MLEs perform more reliably in estimating the hyperparameters.

3.2. The Effect of κ_0 for Quantities of Interest as Functions of x_{n+1}

In this subsection, we will investigate the effect of κ_0 for quantities of interest $(\beta^*, \delta_s^{\pi, \theta_{n+1}}(x_{n+1}), \delta_2^{\pi, \theta_{n+1}}(x_{n+1}), \delta_2^{\pi, \mu_{n+1}}(x_{n+1}), \alpha^*, PESL_s^{\pi, \theta_{n+1}}(x_{n+1}), \text{ and } PESL_2^{\pi, \theta_{n+1}}(x_{n+1}))$ as functions of x_{n+1} . The motivation of this subsection is that κ_0 is unidentifiable, and we want to know whether using $\kappa_0 = 1$ (our recommendation) for the computation will overestimate or underestimate quantities of interest as functions of x_{n+1} when the true $\kappa_0 > 0$ (e.g., 2 or 0.5) is different from $\kappa_0 = 1$.

- Quantities of interest as functions of x_{n+1} for $\kappa_0 = 2, 1, 0.5$ in the simulations are plotted in Figure 2. In all the plots, $u_0 = \sigma_0^2(\kappa_0 + 1)/\kappa_0 = 2$ is fixed. From the figure, we observe the following facts.

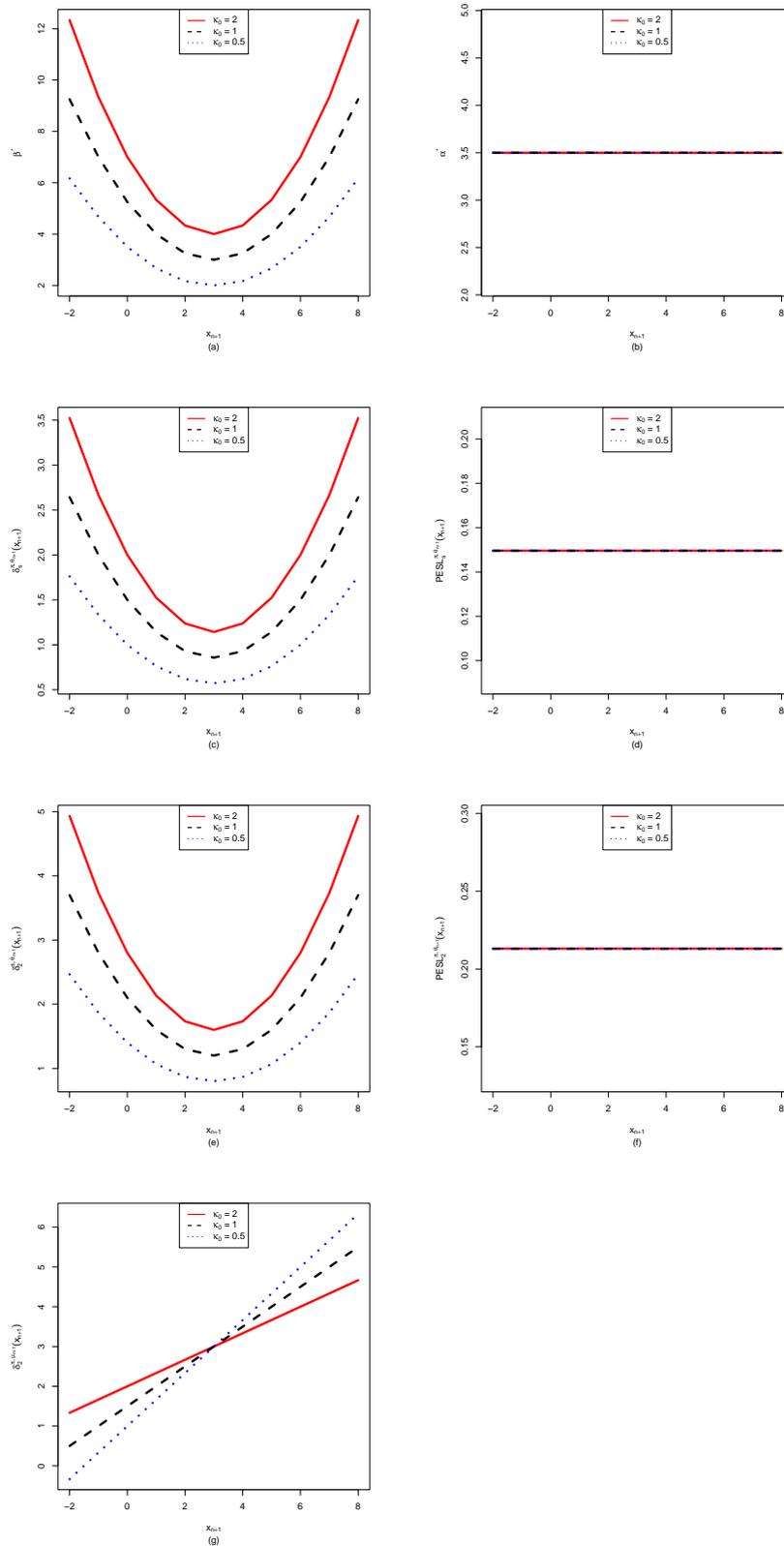


Figure 2. Quantities of interest as functions of x_{n+1} for $\kappa_0 = 2, 1, 0.5$ in the simulations. $u_0 = \sigma_0^2(\kappa_0 + 1)/\kappa_0 = 2$ is fixed in all the plots. Left column: β^* , $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$, $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$, and $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ are affected by κ_0 . Right column: α^* , $PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$, and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$ are not affected by κ_0 .

1. From the left column of the figure, we see that β^* , $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$, $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$, and $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ are affected by κ_0 . Moreover, β^* , $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$, and $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$ are quadratic functions of x_{n+1} (note that their y ranges are different), while $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ is a linear function of x_{n+1} .
2. For plot (a), we observe that β^* when using $\kappa_0 = 1$ will underestimate β^* when using $\kappa_0 = 2$, and it will overestimate β^* when using $\kappa_0 = 0.5$. In other words, β^* is an increasing function of κ_0 for a fixed x_{n+1} . Similar phenomena are observed for $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ (plot (c)) and $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$ (plot (e)).
3. For $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ (plot (g)), we observe an interesting phenomenon that when $x_{n+1} < \mu_0$ ($= 3$ in the simulation), $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ when using $\kappa_0 = 1$ will underestimate that when using $\kappa_0 = 2$, and it will overestimate that when using $\kappa_0 = 0.5$. In other words, $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ is an increasing function of κ_0 when $x_{n+1} < \mu_0$. However, when $x_{n+1} > \mu_0$, we observe a reversed phenomenon that $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ when using $\kappa_0 = 1$ will overestimate that when using $\kappa_0 = 2$, and it will underestimate that when using $\kappa_0 = 0.5$. In other words, $\delta_2^{\pi, \mu_{n+1}}(x_{n+1})$ is a decreasing function of κ_0 when $x_{n+1} > \mu_0$. The phenomena can be explained by the theoretical analysis of

$$\frac{\partial \delta_2^{\pi, \mu_{n+1}}(x_{n+1})}{\partial \kappa_0} = \frac{\mu_0 - x_{n+1}}{(\kappa_0 + 1)^2} \begin{cases} > 0, & \text{if } x_{n+1} < \mu_0, \\ = 0, & \text{if } x_{n+1} = \mu_0, \\ < 0, & \text{if } x_{n+1} > \mu_0. \end{cases}$$

4. From the right column of the figure, we see that α^* , $PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$, and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$ are not affected by κ_0 . Moreover, they also do not depend on x_{n+1} .

3.3. Two Inequalities of the Bayes Estimators and the PESLs

This subsection aims to provide numerical illustrations of the two inequalities related to the Bayes estimators and the PESLs—specifically, (15) and (16)—under the oracle scenario where the hyperparameters $(\mu_0, \kappa_0, \nu_0, \sigma_0^2)$ are known. The rationale for this section lies in the theoretical foundation established by these two inequalities, which we seek to validate through numerical examples.

First, we fix $\mu_0 = 3$, $\kappa_0 = 1$, $\nu_0 = 6$, and $\sigma_0^2 = 1$. Then we set a seed number 1 in R software and draw $\theta_{n+1} = 1.866135$ from $\widetilde{IG}(\nu_0/2, \nu_0\sigma_0^2/2)$. Next, we draw $\mu_{n+1} = 4.816593$ from $N(\mu_0, \theta_{n+1}/\kappa_0)$. After that, we draw $x_{n+1} = 6.554815$ from $N(\mu_{n+1}, \theta_{n+1})$. Figure 3 shows the histogram of $\theta_{n+1}|x_{n+1}$ and the density estimation curve of $\pi(\theta_{n+1}|x_{n+1})$. It is $\pi(\theta_{n+1}|x_{n+1})$ that we find $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ to minimize the PESL. Numerical results show that

$$\delta_s^{\pi, \theta_{n+1}}(x_{n+1}) = 1.759765 < 2.463671 = \delta_2^{\pi, \theta_{n+1}}(x_{n+1})$$

and

$$PESL_s^{\pi, \theta_{n+1}}(x_{n+1}) = 0.1496063 < 0.2131341 = PESL_2^{\pi, \theta_{n+1}}(x_{n+1}),$$

which exemplify the theoretical studies of (15) and (16).

We now consider the scenario where one of the five parameters— μ_0 , κ_0 , ν_0 , σ_0^2 , and x_{n+1} —is allowed to vary while the others remain constant. Specifically, our focus lies on conducting a sensitivity analysis of the Bayes estimators and the PESLs with respect to these five parameters.

Figure 4 illustrates the Bayes estimators and the PESLs as varying with ν_0 . As shown in the left panel, the Bayes estimators exhibit dependence on ν_0 , demonstrating the validity of inequality (15). Specifically, $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ increases as ν_0 increases, whereas $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$ decreases with rising ν_0 . The right panel reveals that the PESLs are also influenced by ν_0 , supporting the inequality (16), and both PESLs decrease monotonically as ν_0 grows. Additionally, numerical values corresponding to these quantities are provided in Table 3, which aligns with the graphical representation in Figure 4. Overall, both the figure and the table confirm the two theoretical inequalities: (15) and (16).

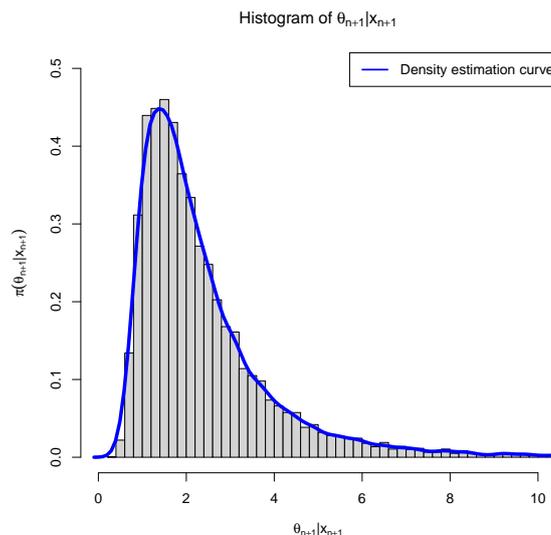


Figure 3. The histogram of $\theta_{n+1}|x_{n+1}$ and the density estimation curve of $\pi(\theta_{n+1}|x_{n+1})$.

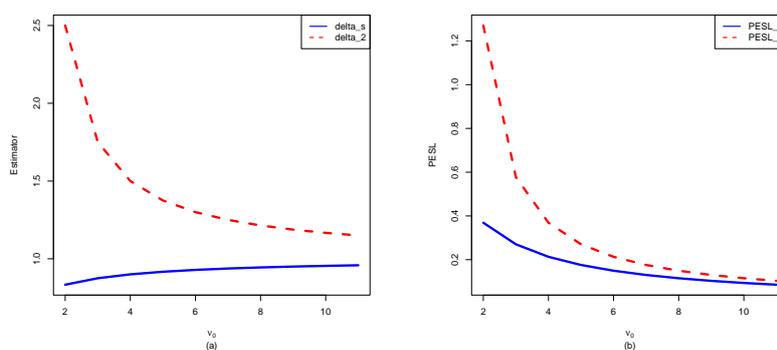


Figure 4. The Bayes estimators and the PESLs as functions of ν_0 . (a) Bayes estimators vs ν_0 . (b) PESLs vs ν_0 .

Table 3. The numerical values of the Bayes estimators and the PESLs in Figure 4: ν_0 changes.

ν_0	2	3	4	5	6	7	8	9	10	11
$\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$	0.8333	0.8750	0.9000	0.9167	0.9286	0.9375	0.9444	0.9500	0.9545	0.9583
$\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$	2.5000	1.7500	1.5000	1.3750	1.3000	1.2500	1.2143	1.1875	1.1667	1.1500
$PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$	0.3690	0.2704	0.2131	0.1758	0.1496	0.1302	0.1152	0.1033	0.0937	0.0856
$PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$	1.2704	0.5772	0.3690	0.2704	0.2131	0.1758	0.1496	0.1302	0.1152	0.1033

Figure 5 shows the Bayes estimators and the PESLs as functions of $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} . We see from the left plots of the figure that the Bayes estimators depend on $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} , and (15) is exemplified. Moreover, the Bayes estimators are first decreasing and then increasing functions of μ_0 and x_{n+1} , and they are increasing functions of κ_0 and σ_0^2 . The right plots of the figure exhibit that the PESLs do not depend on $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} , and (16) is exemplified. Furthermore, Tables 4–7 display the numerical values of the Bayes estimators and the PESLs in Figure 5. In summary, the results of Figure 5 and Tables 4–7 exemplify the two inequalities (15) and (16).

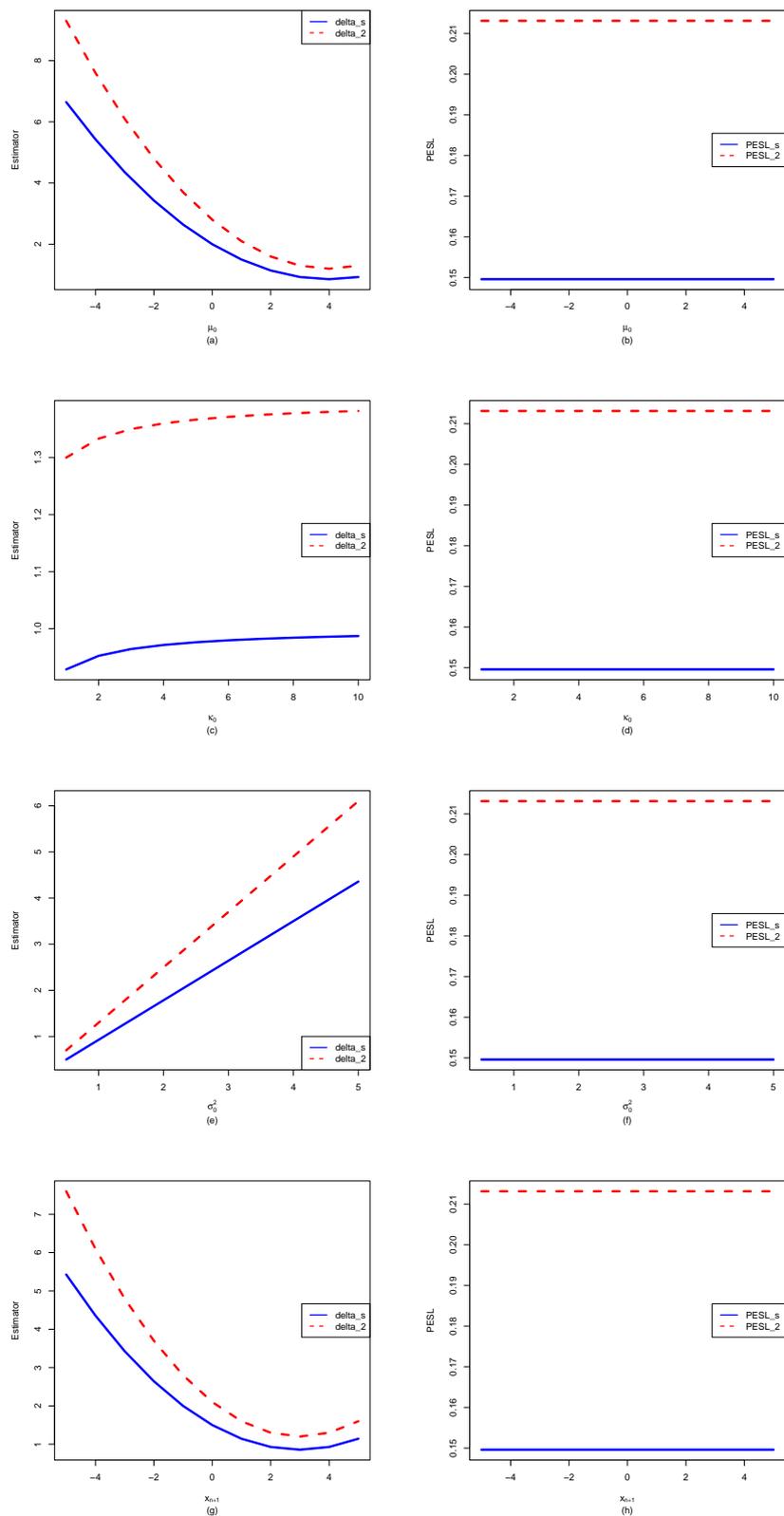


Figure 5. The Bayes estimators and the PESLs as functions of $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} . (a), (c), (e), (g) Bayes estimators vs $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} . (b), (d), (f), (h) PESLs vs $\mu_0, \kappa_0, \sigma_0^2$, and x_{n+1} .

Table 4. The numerical values of the Bayes estimators and the PESLs in Figure 5: μ_0 changes.

μ_0	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\delta_s^{\pi,\theta_{n+1}}(x_{n+1})$	6.6429	5.4286	4.3571	3.4286	2.6429	2.0000	1.5000	1.1429	0.9286	0.8571	0.9286
$\delta_2^{\pi,\theta_{n+1}}(x_{n+1})$	9.3000	7.6000	6.1000	4.8000	3.7000	2.8000	2.1000	1.6000	1.3000	1.2000	1.3000
$PESL_s^{\pi,\theta_{n+1}}(x_{n+1})$	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496
$PESL_2^{\pi,\theta_{n+1}}(x_{n+1})$	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131

Table 5. The numerical values of the Bayes estimators and the PESLs in Figure 5: κ_0 changes.

κ_0	1	2	3	4	5	6	7	8	9	10
$\delta_s^{\pi,\theta_{n+1}}(x_{n+1})$	0.9286	0.9524	0.9643	0.9714	0.9762	0.9796	0.9821	0.9841	0.9857	0.9870
$\delta_2^{\pi,\theta_{n+1}}(x_{n+1})$	1.3000	1.3333	1.3500	1.3600	1.3667	1.3714	1.3750	1.3778	1.3800	1.3818
$PESL_s^{\pi,\theta_{n+1}}(x_{n+1})$	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496
$PESL_2^{\pi,\theta_{n+1}}(x_{n+1})$	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131

Table 6. The numerical values of the Bayes estimators and the PESLs in Figure 5: σ_0^2 changes.

σ_0^2	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
$\delta_s^{\pi,\theta_{n+1}}(x_{n+1})$	0.5000	0.9286	1.3571	1.7857	2.2143	2.6429	3.0714	3.5000	3.9286	4.3571
$\delta_2^{\pi,\theta_{n+1}}(x_{n+1})$	0.7000	1.3000	1.9000	2.5000	3.1000	3.7000	4.3000	4.9000	5.5000	6.1000
$PESL_s^{\pi,\theta_{n+1}}(x_{n+1})$	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496
$PESL_2^{\pi,\theta_{n+1}}(x_{n+1})$	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131

Table 7. The numerical values of the Bayes estimators and the PESLs in Figure 5: x_{n+1} changes.

x_{n+1}	-5	-4	-3	-2	-1	0	1	2	3	4	5
$\delta_s^{\pi,\theta_{n+1}}(x_{n+1})$	5.4286	4.3571	3.4286	2.6429	2.0000	1.5000	1.1429	0.9286	0.8571	0.9286	1.1429
$\delta_2^{\pi,\theta_{n+1}}(x_{n+1})$	7.6000	6.1000	4.8000	3.7000	2.8000	2.1000	1.6000	1.3000	1.2000	1.3000	1.6000
$PESL_s^{\pi,\theta_{n+1}}(x_{n+1})$	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496	0.1496
$PESL_2^{\pi,\theta_{n+1}}(x_{n+1})$	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131	0.2131

In brief, the results of Figures 4 and 5 exemplify that the PESLs depend only on ν_0 , but not on μ_0 , κ_0 , σ_0^2 , and x_{n+1} .

Since the estimators $\delta_2^{\pi,\theta_{n+1}}(x_{n+1})$ and $\delta_s^{\pi,\theta_{n+1}}(x_{n+1})$, along with their corresponding PESLs— $PESL_2^{\pi,\theta_{n+1}}(x_{n+1})$ and $PESL_s^{\pi,\theta_{n+1}}(x_{n+1})$ —are functions of the parameters α^* and β^* , where $\alpha^* > 1$ and $\beta^* > 0$, we visualize their behavior across the domain $D = (1, 10] \times (0, 10]$ using the `persp3d()` function from the R package `rgl` (refer to [20,21,36,37]). A key advantage of `persp3d()` over the standard `persp()` function in the `graphics` package is its ability to overlay multiple surfaces within the same plot and enable interactive rotation of 3D visualizations, thereby offering greater flexibility in graphical analysis. The resulting surface plots for the estimators, the PESLs, and their pairwise differences are displayed in Figure 6. As shown in the two leftmost panels, $\delta_2^{\pi,\theta_{n+1}}(x_{n+1})$ consistently exceeds $\delta_s^{\pi,\theta_{n+1}}(x_{n+1})$ across all values of $(\alpha^*, \beta^*) \in D$, providing empirical support for the inequality stated in (15). Similarly, the rightmost panels indicate that $PESL_2^{\pi,\theta_{n+1}}(x_{n+1})$ remains greater than $PESL_s^{\pi,\theta_{n+1}}(x_{n+1})$ throughout D , which visually confirms the theoretical result in (16). Overall, the numerical illustrations in the figure corroborate the analytical findings presented in (15) and (16).

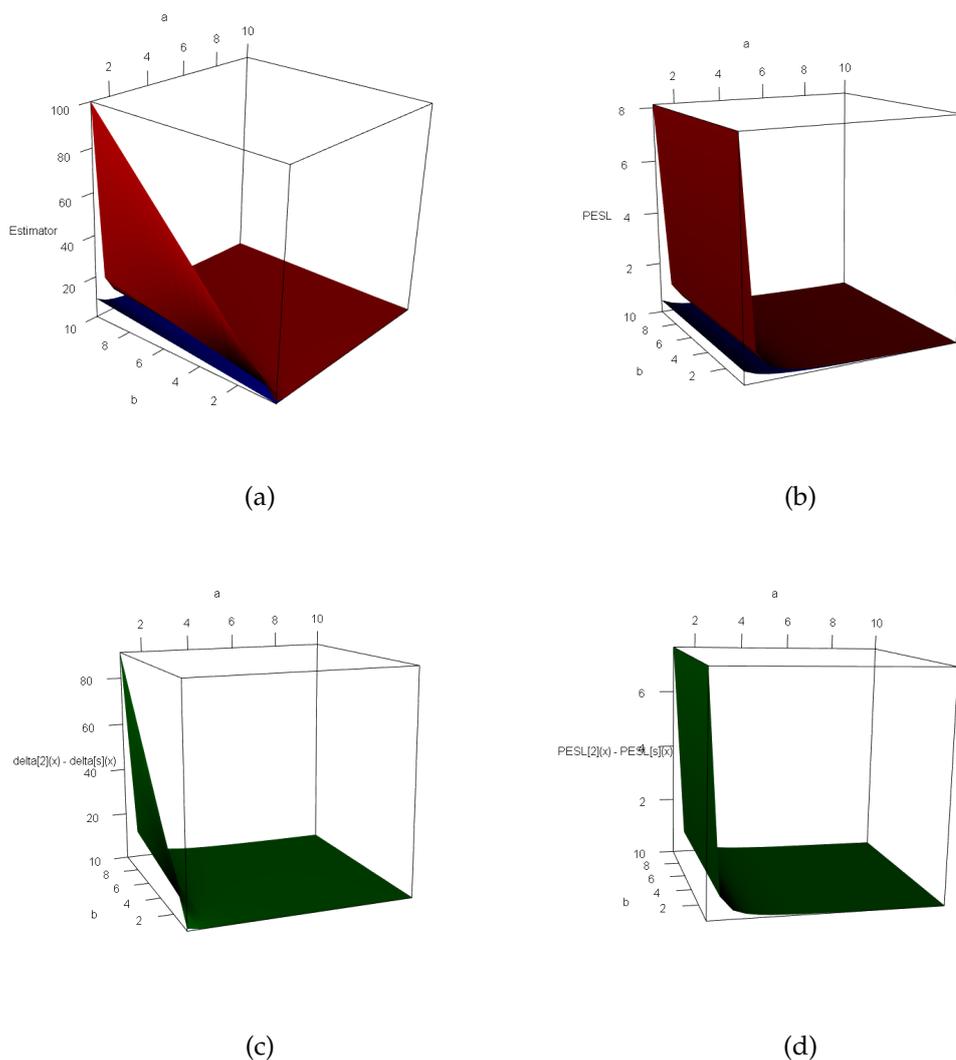


Figure 6. The domain for (α^*, β^*) is $D = (1, 10] \times (0, 10]$ for all the plots. a is for α^* and b is for β^* in the axes of all the plots. The red surface is for $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$ and the blue surface is for $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ in the upper two plots. (a) The estimators as functions of α^* and β^* . $\delta_2^{\pi, \theta_{n+1}}(x_{n+1}) > \delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ for all (α^*, β^*) on D . (b) The PESLs as functions of α^* and β^* . $PESL_2^{\pi, \theta_{n+1}}(x_{n+1}) > PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$ for all (α^*, β^*) on D . (c) The surface of $\delta_2^{\pi, \theta_{n+1}}(x_{n+1}) - \delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ which is positive for all (α^*, β^*) on D . (d) The surface of $PESL_2^{\pi, \theta_{n+1}}(x_{n+1}) - PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$ which is also positive for all (α^*, β^*) on D .

3.4. Goodness-of-Fit of the Model: KS Test

This section focuses on evaluating the goodness-of-fit of the hierarchical normal and normal-inverse-gamma model (2) to the generated synthetic data. The primary objective is to assess how well the marginal distribution of the N-NIG model captures the characteristics of the simulated dataset. It should be noted that the analysis here is based solely on the observed data points $\mathbf{x}^{(n)} = (x_1, x_2, \dots, x_n)$.

The Kolmogorov-Smirnov (KS) statistic measures the maximum discrepancy, denoted as D , between the empirical cumulative distribution function (cdf), $F_n(x)$, and the theoretical population cdf, $F_0(x)$. Mathematically, this is expressed as

$$D = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|. \quad (25)$$

In R, the KS test can be carried out using the built-in function `ks.test()` (refer to [38–40]). While the standard KS test is primarily designed for one-dimensional continuous distributions, various

adaptations known as KS-type tests have also been proposed in the literature for handling discrete data ([41–43]).

The R function `ks.test()` performs the KS test, and its return value is a list containing the components *statistic* (the value of the test statistic, or the D value) and *p.value* (the p-value of the test). In the literature, it is well known that a smaller D value or a larger p-value indicates a better fit of the model to the simulated data. Based on and inspired by the D value and p-value, [21] propose five indices (\bar{D} , $\overline{p - value}$, $(\min D)\%$, $(\max p - value)\%$, $(\text{Accept } H_0)\%$) to compare the three methods, that is, the oracle method, the moment method, and the MLE method in simulations. \bar{D} is the average D values (25) of M simulations (the smaller the better). $\overline{p - value}$ is the average p-values of M simulations (the larger the better). $(\min D)\%$ is the percentage of M simulations which attain the minimum D value in the three methods (the larger the better). The three $(\min D)\%$ values should sum to 1. $(\max p - value)\%$ is the percentage of M simulations which attain the maximum p-value in the three methods (the larger the better). The three $(\max p - value)\%$ values should sum to 1. $(\text{Accept } H_0)\%$ is the percentage of accepting H_0 (defined as p-value > 0.05) in M simulations for each method (the larger the better). Each $(\text{Accept } H_0)\%$ value should be between 0% – 100%.

In our problem, the null hypothesis specifies that

$$H_0 : X \sim N - N - \widetilde{IG}(\mu_0, \nu_0, u_0),$$

where $N - N - \widetilde{IG}(\mu_0, \nu_0, u_0)$ is the marginal distribution of the hierarchical normal and normal-inverse-gamma model (2). The marginal density of the $N - N - \widetilde{IG}(\mu_0, \nu_0, u_0)$ distribution is given by (18) which is obviously one-dimensional and continuous. Therefore, the KS test can serve as an indicator of how well the model fits the data.

It should be noted that the data presented in this section are generated based on the hierarchical normal and normal-inverse-gamma model (2), where the hyperparameters are set to $(\mu_0, \kappa_0, \nu_0, \sigma_0^2) = (3, 1, 6, 1)$ and $u_0 = 2$. Alternative values for these hyperparameters may also be chosen depending on the context.

Table 8 presents the goodness-of-fit results from the KS test for model (2) applied to simulated data. The simulations are based on a sample size of $n = 1,000$ and repeated $M = 100$ times. Hyperparameter estimation for μ_0 , ν_0 , and u_0 is conducted using three distinct approaches. The first is the oracle method, which assumes prior knowledge of the true values: $\mu_0 = 3$, $\nu_0 = 6$, and $u_0 = 2$. The second approach employs the method of moments, utilizing moment estimators derived in Theorem 2. The third approach relies on MLE, where the hyperparameters are numerically computed using the MLEs provided in Theorem 3.

As shown in Table 8, the following observations can be made.

1. The \bar{D} values corresponding to the three approaches are 0.0268, 0.0239, and 0.0204, respectively. This indicates that the MLE method performs best, followed by the moment method, while the oracle method yields the least satisfactory results. A plausible explanation for this outcome is that, as shown in (25), the empirical cdf $F_n(x)$ relies on observed data, and similarly, the theoretical cdf $F_0(x)$ used in both the moment and MLE methods are derived from data. In contrast, the $F_0(x)$ associated with the oracle method does not depend on actual data.
2. The $\overline{p - value}$ corresponding to the three methods are 0.5230, 0.6245, and 0.7674, respectively, indicating that the MLE method performs best, followed by the moment method, with the oracle method ranking last. A plausible explanation for this outcome is provided in the preceding paragraph.
3. The $(\min D)\%$ values for the three approaches are 0.20, 0.16, and 0.64, respectively. For the MLE method, the $(\min D)\%$ value exceeds half of the M simulations. In terms of performance ranking, the methods are ordered as follows: MLE method, oracle method, and moment method.
4. The three methods yield $(\max p - value)\%$ values of 0.20, 0.16, and 0.64, respectively. Since a lower D value is associated with a higher p-value, the method producing the minimum D value will also produce the maximum p-value. As a result, the $(\min D)\%$ and $(\max p - value)\%$ values

- are identical across the three approaches. Based on performance, the methods can be ranked in descending order of preference as follows: MLE method, oracle method, and moment method.
- The (Accept H_0)% values for the three methods are 98%, 99%, and 99%, respectively, indicating that these values are close to 100%. This suggests that all three methods exhibit strong performance in terms of model fit.
 - To summarize, the MLE method consistently achieves the top ranking across all five evaluation metrics— \bar{D} , $\overline{p - \text{value}}$, $(\min D)\%$, $(\max p - \text{value})\%$, and $(\text{Accept } H_0)\%$. When comparing the moment method with the MLE method, the results indicate that the MLE method outperforms the moment method on each of these five criteria.

Table 8. The results of the KS test goodness-of-fit of the model (2) to the simulated data.

	oracle method	moment method	MLE method
\bar{D}	0.0268	0.0239	0.0204
$\overline{p - \text{value}}$	0.5230	0.6245	0.7674
$(\min D)\%$	0.20	0.16	0.64
$(\max p - \text{value})\%$	0.20	0.16	0.64
$(\text{Accept } H_0)\%$	0.98	0.99	0.99

Figure 7 presents the boxplots of the D values and p-values obtained from the three methods. The following observations can be made based on this figure.

- The D values obtained by the oracle method are higher than those of the other two approaches. Since lower D values indicate better performance, the preferred ranking of the three methods is as follows: MLE method, moment method, and then the oracle method.
- The oracle method yields the lowest p-values compared to the other two approaches. Since higher p-values are preferable, the methods can be ranked in descending order of preference as follows: MLE method, moment method, and finally the oracle method.
- Large D values are associated with small p-values, whereas small D values are linked to large p-values.
- Based on the D values and p-values, the MLE method outperforms the method of moments.

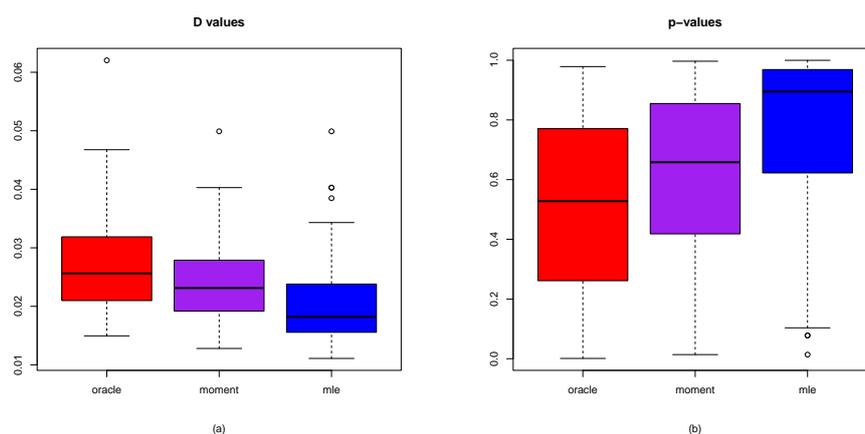


Figure 7. The boxplots of the D values and the p-values for the three methods. (a) D values. (b) p-values.

3.5. Marginal Densities for Various Hyperparameters

This subsection presents the marginal density plots of the hierarchical normal and normal-inverse-gamma model (2) under different settings of the hyperparameters μ_0 , κ_0 , ν_0 , and σ_0^2 . The primary objective is to investigate the types of data distributions that can be captured by this hierarchical model. Recall that the marginal density of X is defined by equation (6), which depends on four hyperparameters: μ_0 (ranging over the real line), $\kappa_0 > 0$, $\nu_0 > 0$, and $\sigma_0^2 > 0$. Our analysis focuses on

how variations in these hyperparameters affect the shape of the marginal density, using the baseline configuration $\mu_0 = 3$, $\kappa_0 = 1$, $\nu_0 = 6$, and $\sigma_0^2 = 1$ as a reference point. Alternative values for the hyperparameters may also be considered to further explore the model's flexibility.

Figure 8 displays the marginal density functions under different values of $\mu_0 = -3, 0, 3, 6, 9$, with $\kappa_0 = 1$, $\nu_0 = 6$, and $\sigma_0^2 = 1$ held constant. It can be observed that as μ_0 increases, the entire distribution shifts rightward without altering its shape, indicating that μ_0 acts as a location parameter. Furthermore, each marginal density is symmetric around its corresponding mean value μ_0 .

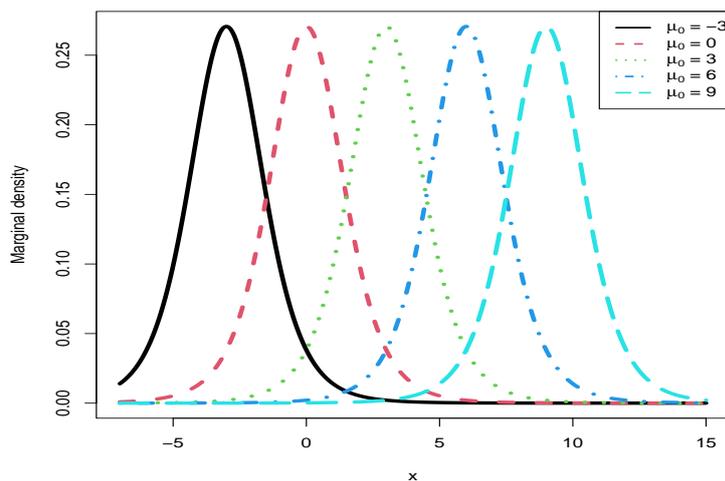


Figure 8. The marginal densities for varied $\mu_0 = -3, 0, 3, 6, 9$, holding $\kappa_0 = 1$, $\nu_0 = 6$, and $\sigma_0^2 = 1$ fixed.

Figure 9 illustrates the marginal density functions under different values of $\kappa_0 = 0.25, 0.5, 1, 2, 4$, with $\mu_0 = 3$, $\nu_0 = 6$, and $\sigma_0^2 = 1$ held constant. It can be observed that as κ_0 grows larger, the peak of the marginal density becomes more pronounced, indicating a reduction in its spread. This is consistent with the theoretical expression for the variance of X , given by

$$\text{Var}(X) = \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0}, \quad (26)$$

which decreases as κ_0 increases. Additionally, each marginal density remains symmetric around the common mean μ_0 .

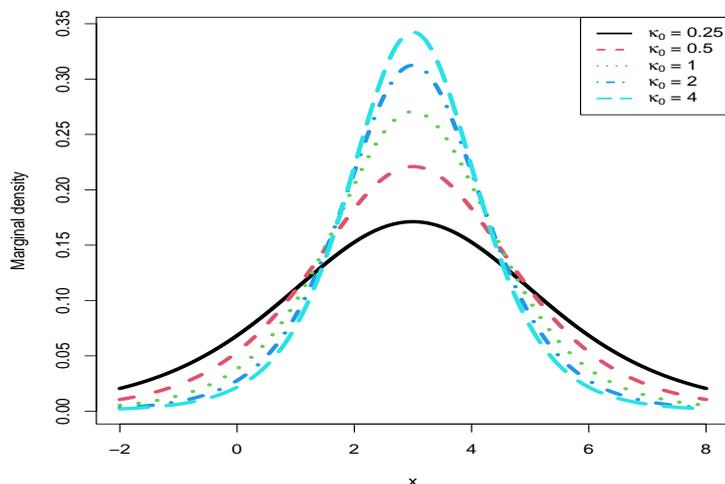


Figure 9. The marginal densities for varied $\kappa_0 = 0.25, 0.5, 1, 2, 4$, holding $\mu_0 = 3$, $\nu_0 = 6$, and $\sigma_0^2 = 1$ fixed.

Figure 10 displays the marginal density functions under different values of $\nu_0 = 2, 4, 6, 12, 24$, with $\mu_0 = 3$, $\kappa_0 = 1$, and $\sigma_0^2 = 1$ held constant. It can be observed that as ν_0 grows larger, the height of the density peak rises, indicating a reduction in the spread of the distribution. This is consistent with (26), which shows that the variance of the marginal density decreases as ν_0 increases. Additionally, each marginal density exhibits symmetry around the common mean μ_0 .

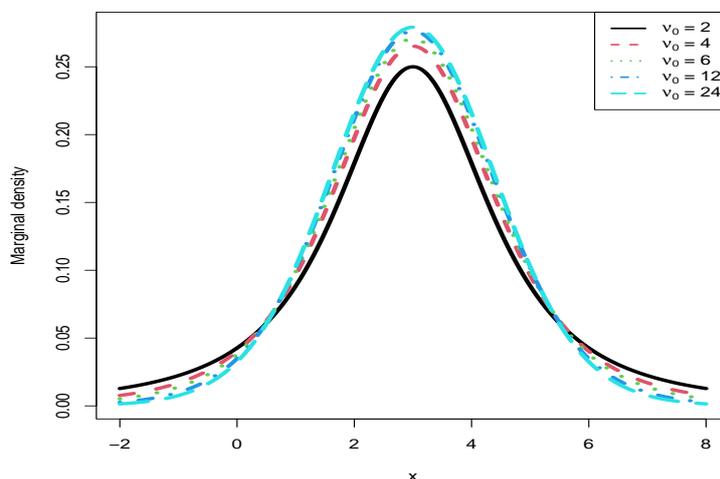


Figure 10. The marginal densities for varied $\nu_0 = 2, 4, 6, 12, 24$, holding $\mu_0 = 3$, $\kappa_0 = 1$, and $\sigma_0^2 = 1$ fixed.

Figure 11 displays the marginal density functions under different values of $\sigma_0^2 = 0.25, 0.5, 1, 2, 4$, with $\mu_0 = 3$, $\kappa_0 = 1$, and $\nu_0 = 6$ held constant. As observed in the plot, an increase in σ_0^2 leads to a reduction in the height of the density peak, indicating a broader distribution. This reflects a larger variance in the marginal density, which aligns with (26) showing that variance increases with σ_0^2 . Additionally, each marginal density remains symmetric around the mean value μ_0 .

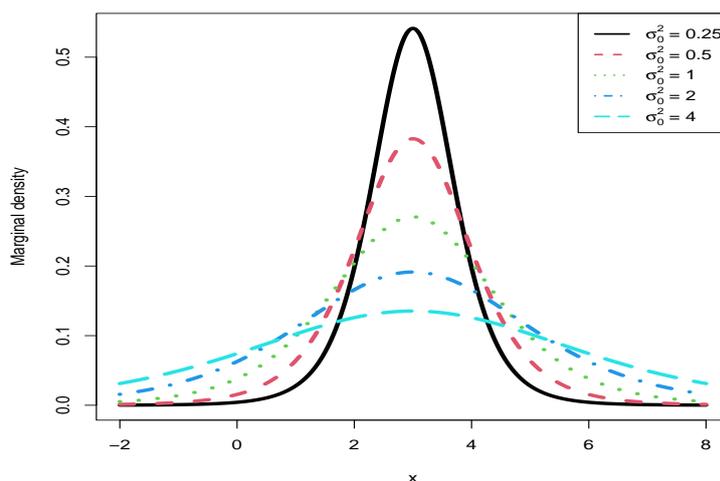


Figure 11. The marginal densities for varied $\sigma_0^2 = 0.25, 0.5, 1, 2, 4$, holding $\mu_0 = 3$, $\kappa_0 = 1$, and $\nu_0 = 6$ fixed.

4. A Real Data Example

In this section, we exploit the poverty level data. The data represent percentages of all persons below the poverty level. The sample is from a random collection of 80 ($= n + 1$) cities in the Western U.S. Source: County and City Data Book 12th edition, U.S. Department of Commerce. The data can be downloaded from

https://college.cengage.com/mathematics/brase/understandable_statistics/7e/students/datasets/svls/frames/frame.html

under section title "Poverty Level". Note that there is an "_" between understandable and statistics.

The histogram of the sample $x^{(n+1)}$ and its density estimation curve is depicted in Figure 12. From the figure, we see that the data are roughly symmetric about 0.15.

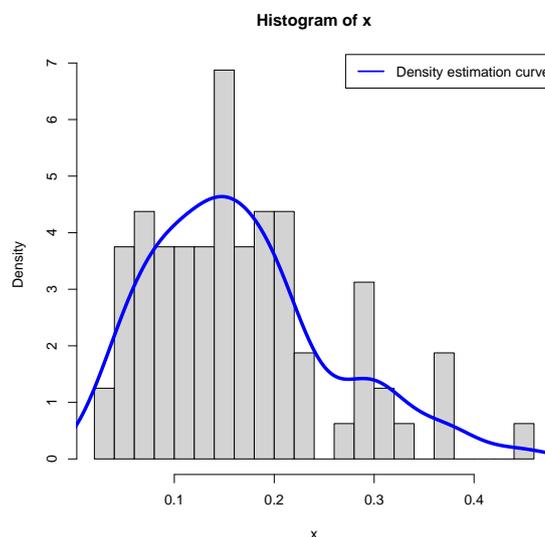


Figure 12. The histogram of the sample $x^{(n+1)}$ and its density estimation curve.

The estimators of the hyperparameters μ_0 , ν_0 , and u_0 , the goodness-of-fit of the model, the empirical Bayes estimators of the variance parameter of the normal distribution with a conjugate normal-inverse-gamma prior and the PESLs, and the mean and variance of the real data by the moment method and the MLE method are summarized in Table 9. From the table, we observe the following facts.

1. The moment estimator of the hyperparameter μ_0 is equal to the sample mean $\bar{x} = 0.1669620$ of the first n observations. It is worth noting that the MLE of the hyperparameter μ_0 is equal to 0.1619865, which is very close to the moment estimator of the hyperparameter μ_0 . Furthermore, the moment estimator and the MLE of the hyperparameter u_0 are also very similar, and they are close to 0.006. However, the moment estimator and the MLE of the hyperparameter ν_0 are quite different. This does not mean that the moment estimator and the MLE are not consistent estimators of the hyperparameter ν_0 , nor mean that the hierarchical normal and normal-inverse-gamma model (2) does not fit the real data. For the big difference between the two estimators, the reason is that the sample size $n = 79$ is small. Certainly, the MLE of the hyperparameter ν_0 is more reliable, as guaranteed from the previous figures and tables in the simulations section.
2. The KS test is used as a measure of the goodness-of-fit. The p-value of the moment method is $0.5313 > 0.05$, which implies that the $t_{\nu_0}(\mu_0, u_0)$ distribution with μ_0 , ν_0 , and u_0 estimated by their moment estimators fits the sample $x^{(n)} = (x_1, x_2, \dots, x_n)$ well. Moreover, the p-value of the MLE method is $0.8233 > 0.05$, which implies that the $t_{\nu_0}(\mu_0, u_0)$ distribution with μ_0 , ν_0 , and u_0 estimated by their MLEs fits the sample $x^{(n)}$ even better. Comparing the moment method and the MLE method, we observe that the D value of the MLEs is smaller, and the p-value of the MLEs is larger, which means that the $t_{\nu_0}(\mu_0, u_0)$ distribution with the hyperparameters μ_0 , ν_0 , and u_0 estimated by the MLEs has a better fit to the sample $x^{(n)}$ than that estimated by the moment estimators.

3. When the hyperparameters are estimated by the MLE method, we see that

$$\delta_s^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)}) = 0.0032576 < 0.0038348 = \delta_2^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)})$$

and

$$PESL_s^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)}) = 0.0771456 < 0.0912061 = PESL_2^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)}).$$

We observe a similar phenomenon for the Bayes estimators and the PESLs when the hyperparameters are estimated by the moment method. Therefore, the two inequalities (15) and (16) are exemplified.

4. The mean of X (the poverty level data) is estimated by $EX \approx \hat{\mu}_0$. The variance of X is estimated by $\text{Var}(X) \approx \frac{\hat{\nu}_0}{\hat{\nu}_0 - 2} \hat{\mu}_0$. It is useful to note that the mean and variance of X by the two methods are very similar, although the estimators of the hyperparameters are very different. Moreover, it is worth mentioning that

$$\begin{aligned} EX &\approx 0.1619865 = 16.19865\%, \\ \text{Var}(X) &\approx 0.0080536 = 80.536\%^2 \end{aligned}$$

for the MLE method. For the moment method, the mean and variance of X are similar. Hence, the variance of X is very small, not large!

Table 9. The estimators of the hyperparameters, the goodness-of-fit of the model, the empirical Bayes estimators of the variance parameter of the normal distribution and the PESLs, and the mean and variance of the real data by the moment method and the MLE method.

		Moment method	MLE method
Estimators of the hyperparameters	μ_0	$\mu_{01}(n) = 0.1669620$	$\mu_{02}(n) = 0.1619865$
	ν_0	$\nu_{01}(n) = 4.933$	$\nu_{02}(n) = 12.287$
	u_0	$u_{01}(n) = 0.005$	$u_{02}(n) = 0.007$
Goodness-of-fit of the model	D	0.0909	0.0708
	p-value	0.5313	0.8233
Empirical Bayes estimators and PESLs	$\delta_s^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)})$	0.0023464	0.0032576
	$\delta_2^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)})$	0.0035396	0.0038348
	$PESL_s^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)})$	0.1779193	0.0771456
	$PESL_2^{\pi, \theta_{n+1}}(\mathbf{x}^{(n+1)})$	0.2753140	0.0912061
Mean and variance of the poverty level data	EX	0.1669620	0.1619865
	Var(X)	0.0080094	0.0080536

5. Conclusions and Discussions

For the hierarchical normal and normal-inverse-gamma (2) framework, we begin by deriving the posterior distributions $\pi(\mu_{n+1}, \theta_{n+1} | x_{n+1})$, $\pi(\theta_{n+1} | x_{n+1})$, $\pi(\mu_{n+1} | x_{n+1})$, and $\pi(\mu_{n+1} | \theta_{n+1}, x_{n+1})$, along with the marginal distribution $m(x_{n+1} | \mu_0, \kappa_0, \nu_0, \sigma_0^2)$, as presented in Theorem 1. Subsequently, we obtain the Bayes estimators $\delta_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $\delta_2^{\pi, \theta_{n+1}}(x_{n+1})$, as well as the corresponding PESLs, denoted by $PESL_s^{\pi, \theta_{n+1}}(x_{n+1})$ and $PESL_2^{\pi, \theta_{n+1}}(x_{n+1})$. These quantities fulfill two key inequalities, given in (15) and (16). Additionally, Theorem 2 provides the moment-based estimates for the model's hyperparameters—specifically $\mu_{01}(n)$, $\nu_{01}(n)$, and $u_{01}(n)$. In parallel, Theorem 3 summarizes the MLEs of the same hyperparameters, namely $\mu_{02}(n)$, $\nu_{02}(n)$, and $u_{02}(n)$. Lastly, Theorem 4 compiles the empirical Bayes estimators for the model's variance component under Stein's loss function, derived using both estimation approaches.

In the simulation study, we conduct numerical experiments on the hierarchical normal and normal-inverse-gamma model (2) from five perspectives: the consistency of moment estimators and

MLEs, the influence of κ_0 on quantities of interest as functions of x_{n+1} , the validation of two inequalities related to Bayes estimators and PESLs, the model's goodness-of-fit to the generated data, and the examination of the model's marginal density distributions.

In the real data analysis, we utilize poverty level data reflecting the proportion of individuals living below the poverty threshold. Table 9 presents a summary of several key results: estimates of the hyperparameters μ_0 , ν_0 , and u_0 ; model goodness-of-fit measures; empirical Bayes estimates for the variance parameter under a normal distribution with a conjugate normal-inverse-gamma prior; PESL values; as well as the mean and variance estimates of the poverty data computed via both the method of moments and MLE. The MLE-based estimates demonstrate a superior fit to the observed sample $x^{(n)}$ compared to moment-based estimates, as evidenced by a smaller D statistic and a higher corresponding p-value, indicating that the $t_{\nu_0}(\mu_0, u_0)$ distribution using MLE-derived hyperparameters better captures the characteristics of the data.

In empirical Bayes approaches, hyperparameters are treated as unknown quantities, and their estimation relies on the marginal distribution derived from observed data. Two widely used techniques for this purpose are the method of moments and MLE. This study applies both methods to estimate the hyperparameters within a hierarchical normal and normal-inverse-gamma model (2).

The simulation results presented in the preceding figures and tables demonstrate that MLEs outperform moment-based estimators in estimating the hyperparameters μ_0 , ν_0 , and u_0 , particularly with respect to consistency and model fit. Nevertheless, these advantages come at a cost. Unlike moment estimators, MLEs are prone to numerical instability, impose higher computational demands, necessitate the constraints $\nu_0 > 0$ and $u_0 > 0$ throughout all iterations, and lack closed-form solutions. Importantly, the performance of MLEs is highly sensitive to the choice of initial values, and moment estimators have been empirically shown to serve as reliable starting points. Furthermore, in scenarios where MLEs fail to converge or do not exist, moment estimators provide a justifiable and robust alternative.

Finally, we outline potential directions for future research. One possible extension involves adapting the hierarchical normal and normal-inverse-gamma model (2) to accommodate various non-conjugate prior distributions for the parameters of the normal distribution (refer to [1,44,45] and related works). In these cases, closed-form solutions may not be feasible, necessitating the development of numerical methods to compute the corresponding estimators.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on [Preprints.org](https://www.preprints.org).

Funding: This study was funded by the National Natural Science Foundation of China (Grant No. 12461051).

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Proofs of Theorems

In the appendix, we will provide some proofs of the theorems.

Appendix A.1. The proof of Theorem 1

Let $\boldsymbol{\eta}_0 = (\mu_0, \kappa_0, \nu_0, \sigma_0^2)'$ be the hyperparameters. The marginal distribution of x_{n+1} in model (2) is

$$m(x_{n+1}|\boldsymbol{\eta}_0) = \frac{f(x_{n+1}|\mu_{n+1}, \theta_{n+1})\pi(\mu_{n+1}, \theta_{n+1}|\boldsymbol{\eta}_0)}{\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1}, \boldsymbol{\eta}_0)}.$$

To lighten notations, the $\boldsymbol{\eta}_0$ will be dropped in the densities. Some of the following derivations are quoted from Example 1.5.1 (p.20) of [25] and [19].

For the random variables, parameters, and hyperparameters, their domains are respectively given by

$$-\infty < x_{n+1}, \mu_{n+1}, \mu_0 < \infty$$

and

$$\theta_{n+1}, \kappa_0, \nu_0, \sigma_0^2 > 0.$$

By the Bayes Theorem, the joint posterior distribution of μ_{n+1} and θ_{n+1} is

$$\pi(\mu_{n+1}, \theta_{n+1} | x_{n+1}) \propto f(x_{n+1} | \mu_{n+1}, \theta_{n+1}) \pi(\mu_{n+1}, \theta_{n+1}).$$

The joint conjugate prior distribution of μ_{n+1} and θ_{n+1} is decomposed as

$$\pi(\mu_{n+1}, \theta_{n+1}) = \pi(\mu_{n+1} | \theta_{n+1}) \pi(\theta_{n+1}),$$

which is a normal-inverse-gamma distribution. Hence,

$$\pi(\mu_{n+1}, \theta_{n+1} | x_{n+1}) \propto f(x_{n+1} | \mu_{n+1}, \theta_{n+1}) \pi(\mu_{n+1} | \theta_{n+1}) \pi(\theta_{n+1}).$$

It is easy to see that

$$\begin{aligned} f(x_{n+1} | \mu_{n+1}, \theta_{n+1}) &= \frac{1}{\sqrt{2\pi} \sqrt{\theta_{n+1}}} \exp \left[-\frac{1}{2\theta_{n+1}} (x_{n+1} - \mu_{n+1})^2 \right] \\ &\propto \theta_{n+1}^{-\frac{1}{2}} \exp \left[-\frac{1}{2\theta_{n+1}} (x_{n+1} - \mu_{n+1})^2 \right], \end{aligned}$$

$$\begin{aligned} \pi(\mu_{n+1} | \theta_{n+1}) &= \frac{1}{\sqrt{2\pi} \sqrt{\theta_{n+1}/\kappa_0}} \exp \left[-\frac{\kappa_0}{2\theta_{n+1}} (\mu_{n+1} - \mu_0)^2 \right] \\ &\propto \theta_{n+1}^{-\frac{1}{2}} \exp \left[-\frac{\kappa_0}{2\theta_{n+1}} (\mu_{n+1} - \mu_0)^2 \right], \end{aligned}$$

and

$$\begin{aligned} \pi(\theta_{n+1}) &= \frac{\left(\frac{\nu_0 \sigma_0^2}{2} \right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \left(\frac{1}{\theta_{n+1}} \right)^{\frac{\nu_0}{2}+1} \exp \left(-\frac{\nu_0 \sigma_0^2}{2} \frac{1}{\theta_{n+1}} \right) \\ &\propto \left(\frac{1}{\theta_{n+1}} \right)^{\frac{\nu_0}{2}+1} \exp \left(-\frac{\nu_0 \sigma_0^2}{2} \frac{1}{\theta_{n+1}} \right). \end{aligned}$$

Therefore,

$$\begin{aligned} &\pi(\mu_{n+1}, \theta_{n+1} | x_{n+1}) \\ &\propto \theta_{n+1}^{-\frac{1}{2}} \exp \left[-\frac{1}{2\theta_{n+1}} (x_{n+1} - \mu_{n+1})^2 \right] \theta_{n+1}^{-\frac{1}{2}} \exp \left[-\frac{\kappa_0}{2\theta_{n+1}} (\mu_{n+1} - \mu_0)^2 \right] \\ &\quad \times \theta_{n+1}^{-\left(\frac{\nu_0}{2}+1\right)} \exp \left(-\frac{\nu_0 \sigma_0^2}{2\theta_{n+1}} \right) \\ &= \theta_{n+1}^{-\frac{1}{2}} \theta_{n+1}^{-\left(\frac{\nu_0+1}{2}+1\right)} \exp \left\{ -\frac{1}{2\theta_{n+1}} \left[(x_{n+1} - \mu_{n+1})^2 + \kappa_0 (\mu_{n+1} - \mu_0)^2 + \nu_0 \sigma_0^2 \right] \right\} \end{aligned} \quad (\text{A.1})$$

The expression in the square brackets of (A.1) changes to

$$\begin{aligned}
 & (x_{n+1} - \mu_{n+1})^2 + \kappa_0(\mu_{n+1} - \mu_0)^2 + \nu_0\sigma_0^2 \\
 &= \mu_{n+1}^2 - 2x_{n+1}\mu_{n+1} + x_{n+1}^2 + \kappa_0\mu_{n+1}^2 - 2\kappa_0\mu_0\mu_{n+1} + \kappa_0\mu_0^2 + \nu_0\sigma_0^2 \\
 &= (\kappa_0 + 1)\mu_{n+1}^2 - 2(x_{n+1} + \kappa_0\mu_0)\mu_{n+1} + x_{n+1}^2 + \kappa_0\mu_0^2 + \nu_0\sigma_0^2 \\
 &= (\kappa_0 + 1)\left[\mu_{n+1}^2 - 2\frac{x_{n+1} + \kappa_0\mu_0}{\kappa_0 + 1}\mu_{n+1}\right] + x_{n+1}^2 + \kappa_0\mu_0^2 + \nu_0\sigma_0^2 \\
 &= (\kappa_0 + 1)\left(\mu_{n+1} - \frac{x_{n+1} + \kappa_0\mu_0}{\kappa_0 + 1}\right)^2 - \frac{(x_{n+1} + \kappa_0\mu_0)^2}{\kappa_0 + 1} + x_{n+1}^2 + \kappa_0\mu_0^2 + \nu_0\sigma_0^2 \\
 &= (\kappa_0 + 1)\left(\mu_{n+1} - \frac{x_{n+1} + \kappa_0\mu_0}{\kappa_0 + 1}\right)^2 + \frac{\kappa_0(x_{n+1} - \mu_0)^2}{\kappa_0 + 1} + \nu_0\sigma_0^2,
 \end{aligned}$$

where

$$\begin{aligned}
 & -\frac{(x_{n+1} + \kappa_0\mu_0)^2}{\kappa_0 + 1} + x_{n+1}^2 + \kappa_0\mu_0^2 \\
 &= \frac{(\kappa_0 + 1)x_{n+1}^2 + (\kappa_0 + 1)\kappa_0\mu_0^2 - (x_{n+1} + \kappa_0\mu_0)^2}{\kappa_0 + 1} \\
 &= \frac{(\kappa_0 + 1)x_{n+1}^2 + (\kappa_0 + 1)\kappa_0\mu_0^2 - x_{n+1}^2 - \kappa_0^2\mu_0^2 - 2\kappa_0\mu_0x_{n+1}}{\kappa_0 + 1} \\
 &= \frac{\kappa_0x_{n+1}^2 + \kappa_0\mu_0^2 - 2\kappa_0\mu_0x_{n+1}}{\kappa_0 + 1} \\
 &= \frac{\kappa_0(x_{n+1}^2 + \mu_0^2 - 2x_{n+1}\mu_0)}{\kappa_0 + 1} \\
 &= \frac{\kappa_0(x_{n+1} - \mu_0)^2}{\kappa_0 + 1}.
 \end{aligned}$$

Let

$$\begin{cases} \mu_1 = (x_{n+1} + \kappa_0\mu_0)/(\kappa_0 + 1), \\ \kappa_1 = \kappa_0 + 1, \\ \nu_1 = \nu_0 + 1, \\ \nu_1\sigma_1^2 = \nu_0\sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1}(x_{n+1} - \mu_0)^2. \end{cases} \quad (\text{A.2})$$

Then (A.1) reduces to

$$\begin{aligned}
 & \pi(\mu_{n+1}, \theta_{n+1}|x_{n+1}) \\
 & \propto \theta_{n+1}^{-\frac{1}{2}} \theta_{n+1}^{-(\frac{\nu_1}{2}+1)} \exp\left\{-\frac{1}{2\theta_{n+1}}\left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\}.
 \end{aligned}$$

It is shown that $\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1})$ is a normal-inverse-gamma distribution as follows. The joint posterior distribution $\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1})$ can be written as $\pi(\mu_{n+1}|\theta_{n+1}, x_{n+1})\pi(\theta_{n+1}|x_{n+1})$, where

$$\begin{cases} \pi(\mu_{n+1}|\theta_{n+1}, x_{n+1}) \propto \left(\frac{\theta_{n+1}}{\kappa_1}\right)^{-1/2} \exp\left[-\frac{\kappa_1}{2\theta_{n+1}}(\mu_{n+1} - \mu_1)^2\right], \\ \pi(\theta_{n+1}|x_{n+1}) \propto \theta_{n+1}^{-(\frac{\nu_1}{2}+1)} \exp\left(-\frac{1}{\theta_{n+1}}\frac{\nu_1\sigma_1^2}{2}\right). \end{cases}$$

That is,

$$\begin{cases} \mu_{n+1}|\theta_{n+1}, x_{n+1} \sim N(\mu_1, \theta_{n+1}/\kappa_1), \\ \theta_{n+1}|x_{n+1} \sim IG(\alpha^* = \nu_1/2, \beta^* = \nu_1\sigma_1^2/2). \end{cases}$$

Therefore, the joint posterior distribution $\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1})$ is

$$\begin{aligned}
& \pi(\mu_{n+1}, \theta_{n+1}|x_{n+1}) \\
&= \pi(\mu_{n+1}|\theta_{n+1}, x_{n+1})\pi(\theta_{n+1}|x_{n+1}) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\theta_{n+1}/\kappa_1}} \exp\left[-\frac{\kappa_1}{2\theta_{n+1}}(\mu_{n+1} - \mu_1)^2\right] \\
&\quad \times \frac{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \theta_{n+1}^{-\left(\frac{\nu_1}{2}+1\right)} \exp\left(-\frac{\nu_1\sigma_1^2}{2} \frac{1}{\theta_{n+1}}\right) \\
&= \frac{\sqrt{\kappa_1}}{\sqrt{2\pi}} \frac{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \theta_{n+1}^{-\frac{1}{2}} \theta_{n+1}^{-\left(\frac{\nu_1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}} \left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\} \\
&\sim N - \widetilde{IG}(\mu_1, \kappa_1, \nu_1, \sigma_1^2).
\end{aligned} \tag{A.3}$$

The joint prior distribution $\pi(\mu_{n+1}, \theta_{n+1})$ is

$$\begin{aligned}
& \pi(\mu_{n+1}, \theta_{n+1}) \\
&= \pi(\mu_{n+1}|\theta_{n+1})\pi(\theta_{n+1}) \\
&= \frac{1}{\sqrt{2\pi}\sqrt{\theta_{n+1}/\kappa_0}} \exp\left[-\frac{\kappa_0}{2\theta_{n+1}}(\mu_{n+1} - \mu_0)^2\right] \\
&\quad \times \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \left(\frac{1}{\theta_{n+1}}\right)^{\frac{\nu_0}{2}+1} \exp\left(-\frac{\nu_0\sigma_0^2}{2} \frac{1}{\theta_{n+1}}\right) \\
&= \frac{\sqrt{\kappa_0}}{\sqrt{2\pi}} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \theta_{n+1}^{-\frac{1}{2}} \theta_{n+1}^{-\left(\frac{\nu_0}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}} \left[\kappa_0(\mu_{n+1} - \mu_0)^2 + \nu_0\sigma_0^2\right]\right\} \\
&\sim N - \widetilde{IG}(\mu_0, \kappa_0, \nu_0, \sigma_0^2).
\end{aligned} \tag{A.4}$$

Comparing (A.4) and (A.3), we find that the normal-inverse-gamma distribution is a conjugate prior for $(\mu_{n+1}, \theta_{n+1})$ of $N(\mu_{n+1}, \theta_{n+1})$.

Now let us derive the marginal posterior density of μ_{n+1} . We have

$$\begin{aligned}
& \pi(\mu_{n+1}|x_{n+1}) \\
&= \int_0^\infty \pi(\mu_{n+1}, \theta_{n+1}|x_{n+1}) d\theta_{n+1} \\
&= \int_0^\infty \frac{\sqrt{\kappa_1}}{\sqrt{2\pi}} \frac{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \theta_{n+1}^{-\left(\frac{\nu_1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}} \left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\} d\theta_{n+1} \\
&= \frac{\sqrt{\kappa_1}}{\sqrt{2\pi}} \frac{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \int_0^\infty \theta_{n+1}^{-\left(\frac{\nu_1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}} \left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\} d\theta_{n+1} \\
&= \frac{\sqrt{\kappa_1}}{\sqrt{2\pi}} \frac{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \frac{\Gamma\left(\frac{\nu_1+1}{2}\right)}{\left[\frac{\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2}{2}\right]^{\frac{\nu_1+1}{2}}},
\end{aligned}$$

by noting that the integrand of the above integral is the kernel of an inverse gamma distribution $\widetilde{IG}(\alpha_1, \beta_1)$ with

$$\alpha_1 = \frac{\nu_1 + 1}{2},$$

$$\beta_1 = \frac{\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2}{2}.$$

Therefore,

$$\begin{aligned} & \pi(\mu_{n+1}|x_{n+1}) \\ &= \frac{\Gamma\left(\frac{\nu_1+1}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)} \frac{\sqrt{\kappa_1}}{\sqrt{2\pi}} \left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}} \frac{1}{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1+1}{2}} \left[1 + \frac{1}{\nu_1} \frac{(\mu_{n+1}-\mu_1)^2}{\sigma_1^2/\kappa_1}\right]^{\frac{\nu_1+1}{2}}} \\ &= \frac{\Gamma\left(\frac{\nu_1+1}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)} \frac{1}{\sqrt{\pi\nu_1\sigma_1^2/\kappa_1}} \left[1 + \frac{1}{\nu_1} \frac{(\mu_{n+1}-\mu_1)^2}{\sigma_1^2/\kappa_1}\right]^{-\frac{\nu_1+1}{2}} \\ &\sim t_{\nu_1}\left(\mu_1, \frac{\sigma_1^2}{\kappa_1}\right). \end{aligned}$$

Now let us calculate

$$\begin{aligned} & f(x_{n+1}|\mu_{n+1}, \theta_{n+1})\pi(\mu_{n+1}, \theta_{n+1}) \\ &= \frac{1}{\sqrt{2\pi}\sqrt{\theta_{n+1}}} \exp\left[-\frac{1}{2\theta_{n+1}}(x_{n+1} - \mu_{n+1})^2\right] \\ &\quad \times \frac{\sqrt{\kappa_0}}{\sqrt{2\pi}} \left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}} \frac{1}{\Gamma\left(\frac{\nu_0}{2}\right)} \theta_{n+1}^{-\left(\frac{\nu_0+1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}}\left[\kappa_0(\mu_{n+1} - \mu_0)^2 + \nu_0\sigma_0^2\right]\right\} \\ &= (2\pi)^{-1} \sqrt{\kappa_0} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \theta_{n+1}^{-\frac{1}{2}} \theta_{n+1}^{-\left(\frac{\nu_0+1}{2}+1\right)} \\ &\quad \times \exp\left\{-\frac{1}{2\theta_{n+1}}\left[(x_{n+1} - \mu_{n+1})^2 + \kappa_0(\mu_{n+1} - \mu_0)^2 + \nu_0\sigma_0^2\right]\right\} \\ &= (2\pi)^{-1} \sqrt{\kappa_0} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \theta_{n+1}^{-\left(\frac{\nu_1+1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}}\left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\}. \end{aligned} \quad (\text{A.5})$$

Finally, let us derive the marginal density of x_{n+1} . Combining (A.5) and (A.3), we have

$$\begin{aligned} & m(x_{n+1}|\mu_0, \kappa_0, \nu_0, \sigma_0^2) \\ &= \frac{f(x_{n+1}|\mu_{n+1}, \theta_{n+1})\pi(\mu_{n+1}, \theta_{n+1})}{\pi(\mu_{n+1}, \theta_{n+1}|x_{n+1})} \\ &= \frac{(2\pi)^{-1} \sqrt{\kappa_0} \frac{\left(\frac{\nu_0\sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\Gamma\left(\frac{\nu_0}{2}\right)} \theta_{n+1}^{-\left(\frac{\nu_1+1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}}\left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\}}{(2\pi)^{-\frac{1}{2}} \sqrt{\kappa_1} \frac{\left(\frac{\nu_1\sigma_1^2}{2}\right)^{\frac{\nu_1}{2}}}{\Gamma\left(\frac{\nu_1}{2}\right)} \theta_{n+1}^{-\left(\frac{\nu_1+1}{2}+1\right)} \exp\left\{-\frac{1}{2\theta_{n+1}}\left[\kappa_1(\mu_{n+1} - \mu_1)^2 + \nu_1\sigma_1^2\right]\right\}} \end{aligned}$$

$$\begin{aligned}
&= \frac{(2\pi)^{-\frac{1}{2}} \sqrt{\kappa_0} \left(\frac{\nu_0 \sigma_0^2}{2}\right)^{\frac{\nu_0}{2}} \Gamma\left(\frac{\nu_0}{2}\right)}{\sqrt{\kappa_1} \left(\frac{\nu_1 \sigma_1^2}{2}\right)^{\frac{\nu_1}{2}} \Gamma\left(\frac{\nu_0}{2}\right)} \\
&= \frac{1}{\sqrt{2\pi}} \frac{\sqrt{\kappa_0}}{\sqrt{\kappa_0 + 1}} \frac{\left(\frac{\nu_0 \sigma_0^2}{2}\right)^{\frac{\nu_0}{2}} \Gamma\left(\frac{\nu_0 + 1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right) \left(\frac{\nu_0 \sigma_0^2 + \frac{\kappa_0}{\kappa_0 + 1} (x_{n+1} - \mu_0)^2}{2}\right)^{\frac{\nu_0 + 1}{2}}} \\
&= \frac{\Gamma\left(\frac{\nu_0 + 1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right) \sqrt{2\pi(\kappa_0 + 1)/\kappa_0}} \frac{\left(\frac{\nu_0 \sigma_0^2}{2}\right)^{\frac{\nu_0}{2}}}{\left(\frac{\nu_0 \sigma_0^2}{2}\right)^{\frac{\nu_0 + 1}{2}}} \left[1 + \frac{1}{\nu_0} \frac{(x_{n+1} - \mu_0)^2}{\sigma_0^2(\kappa_0 + 1)/\kappa_0}\right]^{-\frac{\nu_0 + 1}{2}} \\
&= \frac{\Gamma\left(\frac{\nu_0 + 1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right) \sqrt{\pi \nu_0 \sigma_0^2 (\kappa_0 + 1)/\kappa_0}} \left[1 + \frac{1}{\nu_0} \frac{(x_{n+1} - \mu_0)^2}{\sigma_0^2 (\kappa_0 + 1)/\kappa_0}\right]^{-\frac{\nu_0 + 1}{2}} \\
&\sim t_{\nu_0} \left(\mu_0, \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0}\right).
\end{aligned}$$

The proof of the theorem is complete. \square

Appendix A.2. The Proof of Theorem 2

The hyperparameters of model (2) are $-\infty < \mu_0 < \infty$, $\kappa_0 > 0$, $\nu_0 > 0$, and $\sigma_0 > 0$. By Theorem 1, we know that the marginal distribution of X in model (2) is a non-standardized student-t distribution, that is,

$$X \sim t_{\nu_0} \left(\mu_0, \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0}\right).$$

Since there are four hyperparameters, if we want to obtain the estimators of the hyperparameters of model (2) by the moment method, we need to calculate the first four moments of X at least. By Lemma 4 in [19], we obtain the first six population moments of X as follows. Furthermore, letting the population moments be equal to the sample moments, we obtain

$$\begin{aligned}
EX &= \mu_0 = A_1, \\
EX^2 &= \mu_0^2 + \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} = A_2, \text{ if } \nu_0 > 2, \\
EX^3 &= \mu_0^3 + 3\mu_0 \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} = A_3, \text{ if } \nu_0 > 2, \\
EX^4 &= \mu_0^4 + 6\mu_0^2 \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} + 3 \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} \left[\frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0}\right]^2 = A_4, \text{ if } \nu_0 > 4, \\
EX^5 &= \mu_0^5 + 10\mu_0^3 \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} + 15\mu_0 \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} \left[\frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0}\right]^2 = A_5, \text{ if } \nu_0 > 4, \\
EX^6 &= \left\{ \begin{aligned} &\mu_0^6 + 15\mu_0^4 \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0} + 45\mu_0^2 \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} \left[\frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0}\right]^2 \\ &+ 15 \frac{\nu_0^3}{(\nu_0 - 2)(\nu_0 - 4)(\nu_0 - 6)} \left[\frac{\sigma_0^2 (\kappa_0 + 1)}{\kappa_0}\right]^3 \end{aligned} \right\} = A_6, \text{ if } \nu_0 > 6.
\end{aligned}$$

From the first moment of X , we obtain the moment estimator of μ_0 as

$$\mu_{01}(n) = A_1.$$

Let

$$\begin{aligned} w_1 &= \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0}, \\ w_2 &= \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} \left[\frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0} \right]^2, \\ w_3 &= \frac{\nu_0^3}{(\nu_0 - 2)(\nu_0 - 4)(\nu_0 - 6)} \left[\frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0} \right]^3. \end{aligned}$$

From the second moment of X , we obtain the moment estimator of w_1 as

$$w_1(n) = A_2 - A_1^2.$$

From the third moment of X , we obtain the moment estimator of w_1 as

$$w_1(n) = \frac{A_3 - A_1^3}{3A_1}.$$

Obviously, the two moment estimators of w_1 are not equal. Therefore, we choose one of them as the moment estimator of w_1 . For simplicity, we use the moment estimator of w_1 calculated from EX^2 , and ignore the third equation involving EX^3 . Similarly, the equations involving EX^4 and EX^5 both have w_1 and w_2 . For simplicity, we use the equation involving EX^4 , and ignore the equation involving EX^5 . To have four equations, we will use the equation involving EX^6 . Therefore, the moment equations become

$$\begin{aligned} EX &= \mu_0 = A_1, \\ EX^2 &= \mu_0^2 + w_1 = A_2, \\ EX^4 &= \mu_0^4 + 6\mu_0^2 w_1 + 3w_2 = A_4, \\ EX^6 &= \mu_0^6 + 15\mu_0^4 w_1 + 45\mu_0^2 w_2 + 15w_3 = A_6. \end{aligned}$$

Solving the above moment equations, we obtain

$$\begin{aligned} \mu_{01}(n) &= A_1, \\ w_1(n) &= \frac{\nu_0}{\nu_0 - 2} \frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0} \\ &= A_2 - A_1^2 \triangleq B_1, \\ w_2(n) &= \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} \left[\frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0} \right]^2 \\ &= \frac{1}{3} (A_4 - 6A_1^2 A_2 + 5A_1^4) \triangleq B_2, \\ w_3(n) &= \frac{\nu_0^3}{(\nu_0 - 2)(\nu_0 - 4)(\nu_0 - 6)} \left[\frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0} \right]^3 \\ &= \frac{1}{15} (-61A_1^6 + 75A_2 A_1^4 - 15A_4 A_1^2 + A_6) \triangleq B_3. \end{aligned} \tag{A.6}$$

Let

$$u_0 = \frac{\sigma_0^2(\kappa_0 + 1)}{\kappa_0}.$$

Since κ_0 and σ_0 appear together in u_0 , we can not directly obtain the estimators of κ_0 and σ_0 by the moment method. But we can obtain the estimator of u_0 by the moment method. In the following, we are interested in obtaining the moment estimators of $-\infty < \mu_0 < \infty$, $\nu_0 > 0$, and $u_0 > 0$. The moment equations involving $w_1(n)$, $w_2(n)$, and $w_3(n)$ become

$$\begin{aligned} \frac{\nu_0}{\nu_0 - 2} u_0 &= B_1, \\ \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} u_0^2 &= B_2, \\ \frac{\nu_0^3}{(\nu_0 - 2)(\nu_0 - 4)(\nu_0 - 6)} u_0^3 &= B_3. \end{aligned}$$

Since there are three equations and only two parameters ν_0 and u_0 , for simplicity, we will only use the first two equations, and ignore the third equation. Solving the above first two equations for ν_0 and u_0 , we obtain the moment estimators of ν_0 and u_0 as

$$\begin{aligned} \nu_{01}(n) &= \frac{2B_1^2 - 4B_2}{B_1^2 - B_2}, \\ u_{01}(n) &= \frac{-B_1 B_2}{B_1^2 - 2B_2}. \end{aligned}$$

Substituting the expressions of B_1 and B_2 , and after some algebra, we obtain the expressions of $\nu_{01}(n)$ and $u_{01}(n)$ in terms of A_i ($i = 1, 2, 4$) as

$$\nu_{01}(n) = \frac{-\frac{14}{3}A_1^4 + 4A_1^2A_2 + 2A_2^2 - \frac{4}{3}A_4}{-\frac{2}{3}A_1^4 + A_2^2 - \frac{1}{3}A_4}, \quad (\text{A.7})$$

$$u_{01}(n) = \frac{-\frac{1}{3}(A_2 - A_1^2)(5A_1^4 - 6A_1^2A_2 + A_4)}{-\frac{7}{3}A_1^4 + 2A_1^2A_2 + A_2^2 - \frac{2}{3}A_4}. \quad (\text{A.8})$$

Therefore, the moment estimators of μ_0 , ν_0 , and u_0 are given by (A.6), (A.7), and (A.8).

Now let us show that the moment estimators are consistent estimators of the hyperparameters. It is easy to show that

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{p} EX^k$$

for $k = 1, 2, 4$, where \xrightarrow{p} means convergence in probability. Hence,

$$\begin{aligned} A_1 &\xrightarrow{p} EX = \mu_0, \\ A_2 &\xrightarrow{p} EX^2 = \mu_0^2 + \frac{\nu_0}{\nu_0 - 2} u_0, \\ A_4 &\xrightarrow{p} EX^4 = \mu_0^4 + 6\mu_0^2 \frac{\nu_0}{\nu_0 - 2} u_0 + 3 \frac{\nu_0^2}{(\nu_0 - 2)(\nu_0 - 4)} u_0^2. \end{aligned}$$

Therefore,

$$\mu_{01}(n) = A_1 \triangleq f(A_1) \xrightarrow{p} f(EX) = \mu_0,$$

$$\begin{aligned}
v_{01}(n) &= \frac{-\frac{14}{3}A_1^4 + 4A_1^2A_2 + 2A_2^2 - \frac{4}{3}A_4}{-\frac{2}{3}A_1^4 + A_2^2 - \frac{1}{3}A_4} \\
&\triangleq g(A_1, A_2, A_4) \\
&\xrightarrow{p} g(\text{EX}, \text{EX}^2, \text{EX}^4) \\
&= \frac{-\frac{14}{3}\mu_0^4 + 4\mu_0^2\left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right) + 2\left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right)^2}{-\frac{2}{3}\mu_0^4 + \left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right)^2 - \frac{1}{3}\left(\mu_0^4 + 6\mu_0^2\frac{v_0}{v_0-2}u_0 + 3\frac{v_0^2}{(v_0-2)(v_0-4)}u_0^2\right)} \\
&= v_0,
\end{aligned}$$

and

$$\begin{aligned}
u_{01}(n) &= \frac{-\frac{1}{3}(A_2 - A_1^2)(5A_1^4 - 6A_1^2A_2 + A_4)}{-\frac{7}{3}A_1^4 + 2A_1^2A_2 + A_2^2 - \frac{2}{3}A_4} \\
&\triangleq h(A_1, A_2, A_4) \\
&\xrightarrow{p} h(\text{EX}, \text{EX}^2, \text{EX}^4) \\
&= \frac{-\frac{1}{3}\left(\left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right) - \mu_0^2\right)\left(5\mu_0^4 - 6\mu_0^2\left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right) + \left(\mu_0^4 + 6\mu_0^2\frac{v_0}{v_0-2}u_0 + 3\frac{v_0^2}{(v_0-2)(v_0-4)}u_0^2\right)\right)}{-\frac{7}{3}\mu_0^4 + 2\mu_0^2\left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right) + \left(\mu_0^2 + \frac{v_0}{v_0-2}u_0\right)^2 - \frac{2}{3}\left(\mu_0^4 + 6\mu_0^2\frac{v_0}{v_0-2}u_0 + 3\frac{v_0^2}{(v_0-2)(v_0-4)}u_0^2\right)} \\
&= u_0.
\end{aligned}$$

The proof of the theorem is complete. \square

Appendix A.3. The Proof of Theorem 3

Now we derive the Maximum Likelihood Estimators (MLEs) of μ_0 , v_0 , and u_0 . By Theorem 1, we know that the marginal distribution of X of the model (2) is

$$m(x|\mu_0, v_0, u_0) = \frac{\Gamma\left(\frac{v_0+1}{2}\right)}{\Gamma\left(\frac{v_0}{2}\right)\sqrt{\pi v_0 u_0}} \left[1 + \frac{1}{v_0} \frac{(x - \mu_0)^2}{u_0}\right]^{-\frac{v_0+1}{2}},$$

for $-\infty < x < \infty$, $-\infty < \mu_0 < \infty$, $\nu_0 > 0$, and $u_0 > 0$. Then the likelihood function of μ_0 , ν_0 , and u_0 is

$$\begin{aligned} L(\mu_0, \nu_0, u_0 | \mathbf{x}) &= m(\mathbf{x} | \mu_0, \nu_0, u_0) \\ &= \prod_{i=1}^n m(x_i | \mu_0, \nu_0, u_0) \\ &= \prod_{i=1}^n \frac{\Gamma\left(\frac{\nu_0+1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right) \sqrt{\pi \nu_0 u_0}} \left[1 + \frac{1}{\nu_0} \frac{(x_i - \mu_0)^2}{u_0}\right]^{-\frac{\nu_0+1}{2}} \\ &= \frac{\Gamma^n\left(\frac{\nu_0+1}{2}\right)}{\Gamma^n\left(\frac{\nu_0}{2}\right)} (\pi \nu_0 u_0)^{-\frac{n}{2}} \prod_{i=1}^n \left[1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}\right]^{-\frac{\nu_0+1}{2}}. \end{aligned}$$

Consequently, the log-likelihood function of μ_0 , ν_0 , and u_0 is

$$\begin{aligned} \log L(\mu_0, \nu_0, u_0 | \mathbf{x}) &= n \log \Gamma\left(\frac{\nu_0+1}{2}\right) - n \log \Gamma\left(\frac{\nu_0}{2}\right) - \frac{n}{2} \log(\pi \nu_0 u_0) \\ &\quad + \sum_{i=1}^n \left(-\frac{\nu_0+1}{2}\right) \log \left[1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}\right] \\ &= n \log \Gamma\left(\frac{\nu_0+1}{2}\right) - n \log \Gamma\left(\frac{\nu_0}{2}\right) - \frac{n}{2} (\log \pi + \log \nu_0 + \log u_0) \\ &\quad - \sum_{i=1}^n \frac{\nu_0+1}{2} \log \left[1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}\right]. \end{aligned}$$

Taking partial derivatives with respect to μ_0 , ν_0 , and u_0 and setting them to zeros, we obtain

$$\begin{aligned} \frac{\partial}{\partial \mu_0} \log L &= - \sum_{i=1}^n \frac{\nu_0+1}{2} \frac{1}{1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}} \frac{1}{\nu_0 u_0} 2(x_i - \mu_0)(-1) \\ &= \sum_{i=1}^n \frac{(\nu_0+1)(x_i - \mu_0)}{\nu_0 u_0 + (x_i - \mu_0)^2} \\ &= 0, \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \nu_0} \log L &= n \psi\left(\frac{\nu_0+1}{2}\right) \frac{1}{2} - n \psi\left(\frac{\nu_0}{2}\right) \frac{1}{2} - \frac{n}{2} \frac{1}{\nu_0} \\ &\quad - \frac{1}{2} \sum_{i=1}^n \left\{ 1 \cdot \log \left[1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}\right] + (\nu_0+1) \frac{1}{1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}} \frac{(x_i - \mu_0)^2}{u_0} \left(-\frac{1}{\nu_0^2}\right) \right\} \\ &= \frac{n}{2} \psi\left(\frac{\nu_0+1}{2}\right) - \frac{n}{2} \psi\left(\frac{\nu_0}{2}\right) - \frac{n}{2} \frac{1}{\nu_0} \\ &\quad - \frac{1}{2} \sum_{i=1}^n \left\{ \log \left[1 + \frac{(x_i - \mu_0)^2}{\nu_0 u_0}\right] - \frac{(\nu_0+1)(x_i - \mu_0)^2}{[\nu_0 u_0 + (x_i - \mu_0)^2] \nu_0} \right\} \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} & \frac{\partial}{\partial u_0} \log L \\ &= -\frac{n}{2} \frac{1}{u_0} - \sum_{i=1}^n \frac{v_0 + 1}{2} \frac{1}{1 + \frac{(x_i - \mu_0)^2}{v_0 u_0}} \frac{(x_i - \mu_0)^2}{v_0} \left(-\frac{1}{u_0^2} \right) \\ &= -\frac{n}{2} \frac{1}{u_0} + \frac{1}{2} \sum_{i=1}^n \frac{(v_0 + 1)(x_i - \mu_0)^2}{[v_0 u_0 + (x_i - \mu_0)^2] u_0} \\ &= 0, \end{aligned}$$

where

$$\psi(x) = \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d}{dx} \log \Gamma(x) = \text{digamma}(x)$$

which can be directly calculated in R software by `digamma(x)` ([46]). Let

$$g_i = v_0 u_0 + (x_i - \mu_0)^2.$$

Thus,

$$(x_i - \mu_0)^2 = g_i - v_0 u_0.$$

After some algebra, the above equations reduce to

$$\begin{aligned} f_1(\mu_0, v_0, u_0) &= \frac{1}{n} \sum_{i=1}^n \frac{(v_0 + 1)(x_i - \mu_0)}{v_0 u_0 + (x_i - \mu_0)^2} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{(v_0 + 1)(x_i - \mu_0)}{g_i} \\ &= 0, \end{aligned} \tag{A.9}$$

$$\begin{aligned} f_2(\mu_0, v_0, u_0) &= \frac{1}{2} \psi\left(\frac{v_0 + 1}{2}\right) - \frac{1}{2} \psi\left(\frac{v_0}{2}\right) - \frac{1}{2v_0} \\ &\quad - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \log \left[1 + \frac{(x_i - \mu_0)^2}{v_0 u_0} \right] - \frac{(v_0 + 1)(x_i - \mu_0)^2}{[v_0 u_0 + (x_i - \mu_0)^2] v_0} \right\} \\ &= \frac{1}{2} \psi\left(\frac{v_0 + 1}{2}\right) - \frac{1}{2} \psi\left(\frac{v_0}{2}\right) - \frac{1}{2v_0} \\ &\quad - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \log g_i - \log(v_0 u_0) - \frac{(v_0 + 1)(g_i - v_0 u_0)}{g_i v_0} \right\} \\ &= 0, \end{aligned} \tag{A.10}$$

$$\begin{aligned} f_3(\mu_0, v_0, u_0) &= -\frac{1}{2u_0} + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(v_0 + 1)(x_i - \mu_0)^2}{[v_0 u_0 + (x_i - \mu_0)^2] u_0} \\ &= -\frac{1}{2u_0} + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(v_0 + 1)(g_i - v_0 u_0)}{g_i u_0} \\ &= 0. \end{aligned} \tag{A.11}$$

In (A.9), (A.10), and (A.11), the expressions involving g_i are used for simplifying the R coding.

The MLEs of the hyperparameters μ_0 , ν_0 , and u_0 are the solutions to the equations (A.9), (A.10), and (A.11). The analytical calculations of the MLEs of μ_0 , ν_0 , and u_0 by solving the above equations are impossible, and thus we have to resort to numerical solutions. We can exploit Newton's method to solve the above equations and to obtain the MLEs of μ_0 , ν_0 , and u_0 . The iterative scheme of Newton's method is

$$\mathbf{p}^{(k+1)} = \mathbf{p}^{(k)} - [\mathbf{J}(\mathbf{p}^{(k)})]^{-1} \mathbf{f}(\mathbf{p}^{(k)}), \quad k = 0, 1, \dots,$$

where $\mathbf{J}(\mathbf{p})$ is the Jacobian matrix of $\mathbf{f}(\mathbf{p}) = (f_1(\mathbf{p}), f_2(\mathbf{p}), f_3(\mathbf{p}))'$ and $\mathbf{p} = (\mu_0, \nu_0, u_0)'$. Note that the MLEs of μ_0 , ν_0 , and u_0 are very sensitive to the initial estimators, and the moment estimators are usually proved to be good initial estimators. The Jacobian matrix of μ_0 , ν_0 , and u_0 is given by

$$\mathbf{J} = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial \mu_0} & \frac{\partial f_1}{\partial \nu_0} & \frac{\partial f_1}{\partial u_0} \\ \frac{\partial f_2}{\partial \mu_0} & \frac{\partial f_2}{\partial \nu_0} & \frac{\partial f_2}{\partial u_0} \\ \frac{\partial f_3}{\partial \mu_0} & \frac{\partial f_3}{\partial \nu_0} & \frac{\partial f_3}{\partial u_0} \end{pmatrix},$$

where

$$\begin{aligned} J_{11} &= \frac{\partial f_1}{\partial \mu_0} \\ &= \frac{1}{n} \sum_{i=1}^n (\nu_0 + 1) \frac{(-1) [v_0 u_0 + (x_i - \mu_0)^2] - (x_i - \mu_0) 2(x_i - \mu_0)(-1)}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \\ &= \frac{1}{n} \sum_{i=1}^n (\nu_0 + 1) \frac{2(x_i - \mu_0)^2 - v_0 u_0 - (x_i - \mu_0)^2}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \\ &= \frac{1}{n} \sum_{i=1}^n (\nu_0 + 1) \frac{(x_i - \mu_0)^2 - v_0 u_0}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \\ &= \frac{1}{n} \sum_{i=1}^n (\nu_0 + 1) \frac{g_i - 2v_0 u_0}{g_i^2}, \end{aligned}$$

$$\begin{aligned} J_{12} &= \frac{\partial f_1}{\partial \nu_0} \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0) \frac{1 \cdot [v_0 u_0 + (x_i - \mu_0)^2] - (\nu_0 + 1) u_0}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0) [(x_i - \mu_0)^2 - u_0]}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0)(g_i - v_0 u_0 - u_0)}{g_i^2}, \end{aligned}$$

$$\begin{aligned}
 J_{13} &= \frac{\partial f_1}{\partial u_0} \\
 &= \frac{1}{n} \sum_{i=1}^n (\nu_0 + 1)(x_i - \mu_0)(-1) \frac{1}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \nu_0 \\
 &= -\frac{1}{n} \sum_{i=1}^n \frac{\nu_0(\nu_0 + 1)(x_i - \mu_0)}{[v_0 u_0 + (x_i - \mu_0)^2]^2} \\
 &= -\frac{1}{n} \sum_{i=1}^n \frac{\nu_0(\nu_0 + 1)(x_i - \mu_0)}{g_i^2},
 \end{aligned}$$

$$\begin{aligned}
 J_{21} &= \frac{\partial f_2}{\partial \mu_0} \\
 &= -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \begin{array}{l} \frac{1}{1 + \frac{(x_i - \mu_0)^2}{v_0 u_0}} \frac{1}{v_0 u_0} 2(x_i - \mu_0)(-1) \\ -\frac{\nu_0 + 1}{v_0} \frac{2(x_i - \mu_0)(-1)g_i - (x_i - \mu_0)^2 2(x_i - \mu_0)(-1)}{g_i^2} \end{array} \right\} \\
 &= -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{-2(x_i - \mu_0)}{g_i} + \frac{\nu_0 + 1}{v_0} \frac{2(x_i - \mu_0)g_i - 2(x_i - \mu_0)^3}{g_i^2} \right\} \\
 &= -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{-2(x_i - \mu_0)}{g_i} + \frac{\nu_0 + 1}{v_0} \frac{2(x_i - \mu_0)v_0 u_0}{g_i^2} \right\} \\
 &= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{2(x_i - \mu_0)}{g_i} - \frac{2(\nu_0 + 1)u_0(x_i - \mu_0)}{g_i^2} \right\} \\
 &= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{2(x_i - \mu_0)g_i - 2(x_i - \mu_0)(\nu_0 u_0 + u_0)}{g_i^2} \\
 &= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{2(x_i - \mu_0)[g_i - \nu_0 u_0 - u_0]}{g_i^2} \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0)(g_i - \nu_0 u_0 - u_0)}{g_i^2} \\
 &= J_{12},
 \end{aligned}$$

$$\begin{aligned}
J_{22} &= \frac{\partial f_2}{\partial v_0} \\
&= \frac{1}{2} \psi' \left(\frac{v_0 + 1}{2} \right) \frac{1}{2} - \frac{1}{2} \psi' \left(\frac{v_0}{2} \right) \frac{1}{2} - \frac{1}{2} \left(-\frac{1}{v_0^2} \right) \\
&\quad - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{1 + \frac{(x_i - \mu_0)^2}{v_0 u_0}} \frac{(x_i - \mu_0)^2}{u_0} \left(-\frac{1}{v_0^2} \right) - (x_i - \mu_0)^2 \frac{-\frac{1}{v_0^2} g_i - \left(1 + \frac{1}{v_0}\right) u_0}{g_i^2} \right\} \\
&= \frac{1}{4} \psi' \left(\frac{v_0 + 1}{2} \right) - \frac{1}{4} \psi' \left(\frac{v_0}{2} \right) + \frac{1}{2v_0^2} \\
&\quad + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(x_i - \mu_0)^2}{g_i v_0} - (x_i - \mu_0)^2 \frac{\frac{1}{v_0^2} g_i + \left(1 + \frac{1}{v_0}\right) u_0}{g_i^2} \right\} \\
&= \frac{1}{4} \psi' \left(\frac{v_0 + 1}{2} \right) - \frac{1}{4} \psi' \left(\frac{v_0}{2} \right) + \frac{1}{2v_0^2} \\
&\quad + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(g_i - v_0 u_0) [g_i (v_0 - 1) - u_0 v_0 (v_0 + 1)]}{g_i^2 v_0^2},
\end{aligned}$$

where

$$\begin{aligned}
&\frac{(x_i - \mu_0)^2}{g_i v_0} - (x_i - \mu_0)^2 \frac{\frac{1}{v_0^2} g_i + \left(1 + \frac{1}{v_0}\right) u_0}{g_i^2} \\
&= \frac{g_i - v_0 u_0}{g_i v_0} - \frac{(g_i - v_0 u_0) \left[\frac{1}{v_0^2} g_i + \left(1 + \frac{1}{v_0}\right) u_0 \right]}{g_i^2} \\
&= \frac{g_i - v_0 u_0}{g_i v_0} - \frac{(g_i - v_0 u_0) [g_i + u_0 v_0 (v_0 + 1)]}{g_i^2 v_0^2} \\
&= \frac{(g_i - v_0 u_0) g_i v_0 - (g_i - v_0 u_0) [g_i + u_0 v_0 (v_0 + 1)]}{g_i^2 v_0^2} \\
&= \frac{(g_i - v_0 u_0) [g_i v_0 - g_i - u_0 v_0 (v_0 + 1)]}{g_i^2 v_0^2} \\
&= \frac{(g_i - v_0 u_0) [g_i (v_0 - 1) - u_0 v_0 (v_0 + 1)]}{g_i^2 v_0^2},
\end{aligned}$$

$$\begin{aligned}
J_{23} &= \frac{\partial f_2}{\partial u_0} \\
&= -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{1 + \frac{(x_i - \mu_0)^2}{v_0 u_0}} \frac{(x_i - \mu_0)^2}{v_0} \left(-\frac{1}{u_0^2} \right) \right. \\
&\quad \left. - \frac{(v_0 + 1)(x_i - \mu_0)^2}{v_0} \left(-\frac{1}{g_i^2} \right) v_0 \right\} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(x_i - \mu_0)^2}{g_i u_0} - \frac{(v_0 + 1)(x_i - \mu_0)^2}{g_i^2} \right\} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0)^2 [v_0 u_0 + (x_i - \mu_0)^2] - u_0 (v_0 + 1)(x_i - \mu_0)^2}{g_i^2 u_0}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{v_0 u_0 (x_i - \mu_0)^2 + (x_i - \mu_0)^4 - v_0 u_0 (x_i - \mu_0)^2 - u_0 (x_i - \mu_0)^2}{g_i^2 u_0} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0)^4 - u_0 (x_i - \mu_0)^2}{g_i^2 u_0} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0)^2 [(x_i - \mu_0)^2 - u_0]}{g_i^2 u_0} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(g_i - v_0 u_0)(g_i - v_0 u_0 - u_0)}{g_i^2 u_0},
\end{aligned}$$

$$\begin{aligned}
J_{31} &= \frac{\partial f_3}{\partial \mu_0} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{v_0 + 1}{u_0} \frac{2(x_i - \mu_0)(-1)g_i - (x_i - \mu_0)^2 2(x_i - \mu_0)(-1)}{g_i^2} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{v_0 + 1}{u_0} \frac{2(x_i - \mu_0) [(x_i - \mu_0)^2 - g_i]}{g_i^2} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{v_0 + 1}{u_0} \frac{2(x_i - \mu_0)(-v_0 u_0)}{g_i^2} \\
&= -\frac{1}{n} \sum_{i=1}^n \frac{v_0(v_0 + 1)(x_i - \mu_0)}{g_i^2} \\
&= J_{13},
\end{aligned}$$

$$\begin{aligned}
J_{32} &= \frac{\partial f_3}{\partial v_0} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(x_i - \mu_0)^2}{u_0} \frac{1 \cdot g_i - (v_0 + 1)u_0}{g_i^2} \\
&= \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(g_i - v_0 u_0)(g_i - v_0 u_0 - u_0)}{g_i^2 u_0} \\
&= J_{23},
\end{aligned}$$

$$\begin{aligned}
J_{33} &= \frac{\partial f_3}{\partial u_0} \\
&= -\frac{1}{2} \left(-\frac{1}{u_0^2} \right) + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n (v_0 + 1)(x_i - \mu_0)^2 (-1) \frac{v_0 2u_0 + (x_i - \mu_0)^2}{g_i^2 u_0^2} \\
&= \frac{1}{2u_0^2} - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n (v_0 + 1)(x_i - \mu_0)^2 \frac{v_0 u_0 + g_i}{g_i^2 u_0^2} \\
&= \frac{1}{2u_0^2} - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n (v_0 + 1) \frac{(g_i - v_0 u_0)(g_i + v_0 u_0)}{g_i^2 u_0^2} \\
&= \frac{1}{2u_0^2} - \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \frac{(v_0 + 1)(g_i^2 - v_0^2 u_0^2)}{g_i^2 u_0^2}.
\end{aligned}$$

Note that

$$\psi'(x) = \frac{d^2}{dx^2} \log \Gamma(x) = \text{trigamma}(x)$$

which can be directly calculated in R software by `trigamma(x)` ([46]). In J_{ij} ($i = 1, 2, 3, j = 1, 2, 3$), the expressions involving g_i are used for simplifying the R coding.

Now let us show that the MLEs are consistent estimators of the hyperparameters. From Theorem 10.1.6 in [47], we know that the MLEs are consistent estimators of the hyperparameters under some regularity conditions in Miscellanea 10.6.2 in [47]. The regularity conditions are listed below:

- (C1). We observe X_1, X_2, \dots, X_n , where $X_i \sim m(x|\mu_0, \nu_0, u_0)$ are iid.
- (C2). The parameter is identifiable; that is, if $(\mu_0, \nu_0, u_0) \neq (\mu'_0, \nu'_0, u'_0)$, then $m(x|\mu_0, \nu_0, u_0) \neq m(x|\mu'_0, \nu'_0, u'_0)$.
- (C3). The densities $m(x|\mu_0, \nu_0, u_0)$ have common support, and $m(x|\mu_0, \nu_0, u_0)$ is differentiable in μ_0 , ν_0 , and u_0 .
- (C4). The parameter space Ω contains an open set ω of which the true parameter value $(\mu_{00}, \nu_{00}, u_{00})$ is an interior point.

It remains to show that the marginal distribution $m(x|\mu_0, \nu_0, u_0)$ satisfies all the regularity conditions. Let $\mathcal{X} = (-\infty, \infty)$ be the support set of $m(x|\mu_0, \nu_0, u_0)$.

First, (C1) is satisfied, as X_1, X_2, \dots, X_n is a random sample from $m(x|\mu_0, \nu_0, u_0)$.

Second, let us show that (C2) is satisfied. The parameter is identifiable

$$\begin{aligned} \Leftrightarrow (\mu_0, \nu_0, u_0) \neq (\mu'_0, \nu'_0, u'_0) &\Rightarrow m(x|\mu_0, \nu_0, u_0) \neq m(x|\mu'_0, \nu'_0, u'_0) \text{ for some } x \in \mathcal{X} \\ \Leftrightarrow m(x|\mu_0, \nu_0, u_0) = m(x|\mu'_0, \nu'_0, u'_0) &\text{ for all } x \in \mathcal{X} \Rightarrow (\mu_0, \nu_0, u_0) = (\mu'_0, \nu'_0, u'_0). \end{aligned} \quad (\text{A.12})$$

We have

$$\begin{aligned} m(x|\mu_0, \nu_0, u_0) &= m(x|\mu'_0, \nu'_0, u'_0) \text{ for all } x \in \mathcal{X} \\ \Leftrightarrow \frac{m(x|\mu_0, \nu_0, u_0)}{m(x|\mu'_0, \nu'_0, u'_0)} &= 1 \text{ for all } x \in \mathcal{X}. \end{aligned} \quad (\text{A.13})$$

Note that

$$m(x|\mu_0, \nu_0, u_0) = \frac{\Gamma\left(\frac{\nu_0+1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)\sqrt{\pi\nu_0}u_0} \left[1 + \frac{1}{\nu_0} \frac{(x-\mu_0)^2}{u_0}\right]^{-\frac{\nu_0+1}{2}},$$

and thus

$$m(x|\mu'_0, \nu'_0, u'_0) = \frac{\Gamma\left(\frac{\nu'_0+1}{2}\right)}{\Gamma\left(\frac{\nu'_0}{2}\right)\sqrt{\pi\nu'_0}u'_0} \left[1 + \frac{1}{\nu'_0} \frac{(x-\mu'_0)^2}{u'_0}\right]^{-\frac{\nu'_0+1}{2}}.$$

Therefore, (A.13) is equivalent to

$$\begin{aligned} \frac{\Gamma\left(\frac{\nu_0+1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)\sqrt{\pi\nu_0}u_0} \frac{\Gamma\left(\frac{\nu'_0}{2}\right)\sqrt{\pi\nu'_0}u'_0}{\Gamma\left(\frac{\nu'_0+1}{2}\right)} \frac{\left[1 + \frac{1}{\nu_0} \frac{(x-\mu_0)^2}{u_0}\right]^{-\frac{\nu_0+1}{2}}}{\left[1 + \frac{1}{\nu'_0} \frac{(x-\mu'_0)^2}{u'_0}\right]^{-\frac{\nu'_0+1}{2}}} &= 1 \text{ for all } x \in \mathcal{X} \\ \Leftrightarrow \frac{\Gamma\left(\frac{\nu_0+1}{2}\right)\Gamma\left(\frac{\nu'_0}{2}\right)\sqrt{\pi\nu'_0}u'_0}{\Gamma\left(\frac{\nu'_0+1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)\sqrt{\pi\nu_0}u_0} \frac{\left[1 + \frac{1}{\nu'_0} \frac{(x-\mu'_0)^2}{u'_0}\right]^{\frac{\nu'_0+1}{2}}}{\left[1 + \frac{1}{\nu_0} \frac{(x-\mu_0)^2}{u_0}\right]^{\frac{\nu_0+1}{2}}} &= 1 \text{ for all } x \in \mathcal{X}, \end{aligned}$$

which implies

$$\frac{\Gamma\left(\frac{\nu_0+1}{2}\right)\Gamma\left(\frac{\nu'_0}{2}\right)\sqrt{\pi\nu'_0u'_0}}{\Gamma\left(\frac{\nu'_0+1}{2}\right)\Gamma\left(\frac{\nu_0}{2}\right)\sqrt{\pi\nu_0u_0}} = \frac{1}{c_1},$$

$$\frac{\left[1 + \frac{1}{\nu'_0} \frac{(x-\mu'_0)^2}{u'_0}\right]^{\frac{\nu'_0+1}{2}}}{\left[1 + \frac{1}{\nu_0} \frac{(x-\mu_0)^2}{u_0}\right]^{\frac{\nu_0+1}{2}}} = c_1 \text{ for all } x \in \mathcal{X}, \quad (\text{A.14})$$

where c_1 is a constant which does not depend on x but may depend on $\mu_0, \nu_0, u_0, \mu'_0, \nu'_0$, and u'_0 . From (A.14), we obtain

$$\mu_0 = \mu'_0 \quad (\text{A.15})$$

and

$$\nu_0u_0 = \nu'_0u'_0. \quad (\text{A.16})$$

Substituting (A.15) and (A.16) into (A.14), we obtain

$$\left[1 + \frac{(x-\mu_0)^2}{\nu_0u_0}\right]^{\frac{\nu'_0-\nu_0}{2}} = c_1 \text{ for all } x \in \mathcal{X},$$

which implies

$$\nu_0 = \nu'_0. \quad (\text{A.17})$$

Substituting (A.17) into (A.16), we obtain

$$u_0 = u'_0. \quad (\text{A.18})$$

Therefore,

$$(\mu_0, \nu_0, u_0) = (\mu'_0, \nu'_0, u'_0).$$

Consequently, (A.12) is correct, and (C2) is satisfied.

Third, (C3) is satisfied, as the densities $m(x|\mu_0, \nu_0, u_0)$ have common support $\mathcal{X} = (-\infty, \infty)$, and $m(x|\mu_0, \nu_0, u_0)$ is differentiable in μ_0, ν_0 , and u_0 .

Finally, (C4) is satisfied, as the true parameter value

$$(\mu_{00}, \nu_{00}, u_{00}) \in \omega = \Omega = (-\infty, \infty) \times (0, \infty) \times (0, \infty),$$

which is an open set.

Therefore, the marginal densities $m(x|\mu_0, \nu_0, u_0)$ satisfy all the regularity conditions, and the MLEs are consistent estimators of the hyperparameters.

The proof of the theorem is complete. \square

References

1. Berger JO. *Statistical Decision Theory and Bayesian Analysis*. Springer: New York, 1985. 2nd edition.
2. Maritz JS, Lwin T. *Empirical Bayes Methods*. Chapman & Hall: London, 1989. 2nd edition.
3. Carlin BP, Louis A. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall: London, 2000. 2nd edition.
4. Robbins H. An empirical bayes approach to statistics. *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, University of California Press, 1955.
5. Robbins H. The empirical bayes approach to statistical decision problems. *Annals of Mathematical Statistics* 1964; 35:1–20.
6. Robbins H. Some thoughts on empirical bayes estimation. *Annals of Statistics* 1983; 1:713–723.

7. Deely JJ, Lindley DV. Bayes empirical bayes. *Journal of the American Statistical Association* 1981; **76**:833–841.
8. Morris C. Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association* 1983; **78**:47–65.
9. Maritz JS, Lwin T. Assessing the performance of empirical bayes estimators. *Annals of the Institute of Statistical Mathematics* 1992; **44**:641–657.
10. Carlin BP, Louis A. Empirical bayes: Past, present and future. *Journal of the American Statistical Association* 2000; **95**:1286–1290.
11. Pensky M. Locally adaptive wavelet empirical bayes estimation of a location parameter. *Annals of the Institute of Statistical Mathematics* 2002; **54**:83–99.
12. Coram M, Tang H. Improving population-specific allele frequency estimates by adapting supplemental data: An empirical bayes approach. *The Annals of Applied Statistics* 2007; **1**:459–479.
13. Efron B. Tweedie's formula and selection bias. *Journal of the American Statistical Association* 2011; **106**:1602–1614.
14. van Houwelingen HC. The role of empirical bayes methodology as a leading principle in modern medical statistics. *Biometrical Journal* 2014; **56**:919–932.
15. Ghosh M, Kubokawa T, Kawakubo Y. Benchmarked empirical bayes methods in multiplicative area-level models with risk evaluation. *Biometrika* 2015; **102**:647–659.
16. Satagopan JM, Sen A, Zhou Q, Lan Q, Rothman N, Langseth H, Engel LS. Bayes and empirical bayes methods for reduced rank regression models in matched case-control studies. *Biometrics* 2016; **72**:584–595.
17. Martin R, Mess R, Walker SG. Empirical bayes posterior concentration in sparse high-dimensional linear models. *Bernoulli* 2017; **23**:1822–1847.
18. Mikulich-Gilbertson SK, Wagner BD, Grunwald GK, Riggs PD, Zerbe GO. Using empirical bayes predictors from generalized linear mixed models to test and visualize associations among longitudinal outcomes. *Statistical Methods in Medical Research* 2018; DOI: 10.1177/0962280218758357.
19. Zhang YY, Rong TZ, Li MM. The empirical bayes estimators of the mean and variance parameters of the normal distribution with a conjugate normal-inverse-gamma prior by the moment method and the mle method. *Communications in Statistics-Theory and Methods* 2019; **48**:2286–2304.
20. Zhang YY, Wang ZY, Duan ZM, Mi W. The empirical bayes estimators of the parameter of the poisson distribution with a conjugate gamma prior under stein's loss function. *Journal of Statistical Computation and Simulation* 2019; **89**:3061–3074.
21. Sun J, Zhang YY, Sun Y. The empirical bayes estimators of the rate parameter of the inverse gamma distribution with a conjugate inverse gamma prior under stein's loss function. *Journal of Statistical Computation and Simulation* 2021; **91**:1504–1523.
22. Zhou MQ, Zhang YY, Sun Y, Sun J, Rong TZ, Li MM. The empirical bayes estimators of the probability parameter of the beta-negative binomial model under zhang's loss function. *Chinese Journal of Applied Probability and Statistics* 2021; **37**:478–494.
23. Sun Y, Zhang YY, Sun J. The empirical bayes estimators of the parameter of the uniform distribution with an inverse gamma prior under stein's loss function. *Communications in Statistics-Simulation and Computation* 2024; DOI: 10.1080/03610918.2022.2093904.
24. Zhang YY, Zhang YY, Wang ZY, Sun Y, Sun J. The empirical bayes estimators of the variance parameter of the normal distribution with a conjugate inverse gamma prior under stein's loss function. *Communications in Statistics-Theory and Methods* 2024; **53**:170–200.
25. Mao SS, Tang YC. *Bayesian Statistics*. China Statistics Press: Beijing, 2012. 2nd edition.
26. Chen MH. *Bayesian Statistics Lecture*. Statistics Graduate Summer School, School of Mathematics and Statistics, Northeast Normal University, Changchun, China July 2014.
27. James W, Stein C. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1961; **1**:361–380.
28. Zhang YY. The bayes rule of the variance parameter of the hierarchical normal and inverse gamma model under stein's loss. *Communications in Statistics-Theory and Methods* 2017; **46**:7125–7133.
29. Xie YH, Song WH, Zhou MQ, Zhang YY. The bayes posterior estimator of the variance parameter of the normal distribution with a normal-inverse-gamma prior under stein's loss. *Chinese Journal of Applied Probability and Statistics* 2018; **34**:551–564.
30. Zhang YY, Xie YH, Song WH, Zhou MQ. Three strings of inequalities among six bayes estimators. *Communications in Statistics-Theory and Methods* 2018; **47**:1953–1961.

31. Good IJ. Turing's anticipation of empirical bayes in connection with the cryptanalysis of the naval enigma. *Journal of Statistical Computation and Simulation* 2000; **66**(2):101–111.
32. Noma H, Matsui S. Empirical bayes ranking and selection methods via semiparametric hierarchical mixture models in microarray studies. *Statistics in Medicine* 2013; **32**(11):1904–1916.
33. Pan W, Jeong KS, Xie Y, Khodursky A. A nonparametric empirical bayes approach to joint modeling of multiple sources of genomic data. *Statistica Sinica* 2008; **18**(2):709–729.
34. Zhang Q, Xu Z, Lai Y. An empirical bayes approach for the identification of long-range chromosomal interaction from hi-c data. *Statistical Applications in Genetics and Molecular Biology* 2021; **20**(1):1–15.
35. Soloff JA, Guntuboyina A, Sen B. Multivariate, heteroscedastic empirical bayes via nonparametric maximum likelihood. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 2024; **87**(1):1–32.
36. Adler D, Murdoch D, others. *rgl: 3D Visualization Using OpenGL* 2017. URL <https://CRAN.R-project.org/package=rgl>, r package version 0.98.1.
37. Zhang YY, Zhou MQ, Xie YH, Song WH. The bayes rule of the parameter in (0,1) under the power-log loss function with an application to the beta-binomial model. *Journal of Statistical Computation and Simulation* 2017; **87**:2724–2737.
38. Conover WJ. Pages 295–301 (one-sample kolmogorov test), 309–314 (two-sample smirnov test). *Practical Nonparametric Statistics*, John Wiley & Sons: New York, 1971.
39. Durbin J. *Distribution Theory for Tests Based on the Sample Distribution Function*. SIAM: Philadelphia, 1973.
40. Marsaglia G, Tsang WW, Wang JB. Evaluating kolmogorov's distribution. *Journal of Statistical Software* 2003; **8**(18).
41. Dimitrova DS, Kaishev VK, Tan S. Computing the kolmogorov-smirnov distribution when the underlying cdf is purely discrete, mixed or continuous 2017; City, University of London Institutional Repository.
42. Aldirawi H, Yang J, Metwally AA. Identifying appropriate probabilistic models for sparse discrete omics data. *IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, IEEE, 2019.
43. Santitissadeekorn N, Lloyd DJB, Short MB, Delahaies S. Approximate filtering of conditional intensity process for poisson count data: Application to urban crime. *Computational Statistics & Data Analysis* 2020; **144**:1–14. Document Number 106850.
44. Berger JO. The case for objective bayesian analysis. *Bayesian Analysis* 2006; **1**:385–402.
45. Berger JO, Bernardo JM, Sun DC. Overall objective priors. *Bayesian Analysis* 2015; **10**:189–221.
46. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria 2024. URL <http://www.R-project.org/>.
47. Casella G, Berger RL. *Statistical Inference*. Duxbury: Pacific Grove, 2002. 2nd edition.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.