axioms

MDPI

Article

Bayesian Inference for a Hidden Truncated Bivariate Exponential Distribution with Applications

Indranil Ghosh [1], Hon Keung Tony Ng [2,*], Kipum Kim [3] and Seong W. Kim [3,*]

[1] Department of Mathematics and Statistics, University of North Carolina, Wilmington, NC 28403, USA; ghoshi@uncw.edu
[2] Department of Mathematical Sciences, Bentley University, Waltham, MA 02452, USA
[3] Department of Mathematical Data Sciences, Hanyang University, Ansan 15588, Republic of Korea; rainbowlion@hanyang.ac.kr
* Correspondence: tng@bentley.edu (H.K.T.N.); seong@hanyang.ac.kr (S.W.K.); Tel.: +1-781-216-7102 (H.K.T.N.); +82-10-400-5467 (S.W.K.)

Abstract: In many real-life scenarios, one variable is observed only if the other concomitant variable or the set of concomitant variables (in the multivariate scenario) is truncated from below, above, or from a two-sided approach. Hidden truncation models have been applied to analyze data when bivariate or multivariate observations are subject to some form of truncation. While the statistical inference for hidden truncation models (truncation from above) under the frequentist and the Bayesian paradigms has been adequately discussed in the literature, the estimation of a two-sided hidden truncation model under the Bayesian framework has not yet been discussed. In this paper, we consider the Bayesian inference for a general two-sided hidden truncation model based on the Arnold–Strauss bivariate exponential distribution. In addition, a Bayesian model selection approach based on the Bayes factor to select between models without truncation, with truncation from below, from above, and two-sided truncation is also explored. An extensive simulation study is carried out for varying parameter choices under the conjugate prior set-up. For illustrative purposes, a real-life dataset is re-analyzed to demonstrate the applicability of the proposed methodology.

Keywords: Bayes factor; bivariate exponential distribution; Gibbs sampling; hidden truncation; informative prior; posterior probability

MSC: 62F15; 60E05; 62F10

Citation: Ghosh, I.; Ng, H.K.T.; Kim, K.; Kim, S.W. Bayesian Inference for a Hidden Truncated Bivariate Exponential Distribution with Applications. *Axioms* **2024**, *13*, 140. https://doi.org/10.3390/axioms13030140

Academic Editors: Eva T. López Sanjuán and María Isabel Parra Arévalo

Received: 13 January 2024
Revised: 4 February 2024
Accepted: 16 February 2024
Published: 22 February 2024

1. Introduction

A hidden truncation model, also known as a selective reporting model or latent variable model, is a mathematical model that describes observed variables truncated with respect to some hidden covariable(s). An example of hidden truncation provided by Arnold and Beaver [1] is that the observed variable is the waist size for uniforms of elite troops, and the hidden covariable is the height of the troops. The elite troops are selected only if they meet a specific minimum height requirement. Therefore, the variable waist size will not be observed unless the troop meets the minimum height requirement. Hidden truncation models have been used in studying personal income data when an individual's income is either not always correctly specified or may be unreported. Different hidden truncation models and their inference are developed mostly using the classical approach. If the data are subject to hidden truncation, then it is natural to expect that the behavior of the random phenomenon differs from a non-truncated model. Note that the notion of the mixture model is related to the hidden truncated model. Arnold and Gomez [2] showed that the two-parameter skew–normal distribution can be viewed as having arisen from a standard normal density by hidden truncation or an additive component construction, and they conjectured that this apparent relation between hidden truncation and mixture models

only occurs in the normal case. However, the hidden truncation and mixture models are slightly different, as in a mixture model, there is no such hidden covariable subject to truncation. Sometimes, through a natural mechanism or otherwise, a dataset has already been subject to one of the types of hidden truncation, such as the Quasar data as originally and independently studied in Efron and Petrosian [3].

A non-exhaustive list of the references on hidden truncation models is as follows. One may cite the seminal paper by Azzalini [4] to propel research endeavors in this area. It is conjectured and supported by appropriate real-life scenarios, including but not limited to income modeling, astronomical data, survival analysis, etc., such that hidden truncation models provide flexible alternatives to the somewhat constrained and elliptically contoured distributions for modeling bivariate/multivariate or matrix-variate data in higher dimensions. Arnold [5] developed and studied several univariate, bivariate, and multivariate parametric families of distributions based on hidden truncation. Arnold and Beaver [1] suggested that a hidden truncation model involves a plethora of model parameters. Subsequently, it almost always involves a ubiquitous normalizing constant, making inferences under a frequentist framework virtually impossible. This is mainly because parameter estimates in conjunction with such models will not be available in closed forms or at least analytically intractable, and quite often, they have constraints. Furthermore, using a frequentist approach, non-normal models involving a two-sided truncation would require special attention when exploring inferential aspects. However, efforts have been made to estimate the model parameters for hidden truncation models under the frequentist approach with some limitations. For example, Ghosh and Nadarajah [6] discussed the inference under the classical approach of a hidden truncated Pareto (type-II) model starting from a bivariate Pareto (type-II) model. Hidden truncation in bivariate and multivariate Pareto data has been developed and applied to income modeling (see the references in [7]). Zaninetti [8] found that a left-truncated beta distribution better fits the initial mass function for stars compared to the lognormal distribution, which has commonly been used in astrophysics. Kotz et al. [9] (Chapter 44) also discussed the details of the formation of distributions through truncation.

Under the Bayesian paradigm, one-sided hidden truncated Pareto (type-II and type-IV) has been discussed by Ghosh [10,11], but not much discussion has been made on the applicability as well as the efficacy of the proposed Bayesian methodology in terms of prior choices, including the choice of hyperparameters, among others. Noticeably, nonparametric methods for estimation corresponding to two-sided truncated data have been developed by Efron and Petrosian [3]. This serves as a major motivation for the current research work.

In this paper, we explore the estimation of model parameters for a two-sided hidden truncated model, assuming that the data come from a bivariate exponential distribution defined by Arnold and Strauss [12]. Since the exponential distribution is an important probability model for studies involving non-negative random variables, such as the frailty models, this probability model is selected to utilize the proposed Bayesian estimation strategy with the hope that the proposed technique can be well adopted for several other non-normal models. For a detailed study on the genesis of the construction of hidden truncated non-normal models, readers may refer to [7] and the references therein. Next, we delve into the hidden truncation models and discuss the reason to start with a bivariate exponential-type model. Hidden truncation involves ubiquitous normalizing constants followed by the truncation parameter(s) masking within other parameters embedded in such a way that there is an estimate for the composite function, say $\psi(\theta)$, in which θ involves not only the truncation parameter but also location and scale parameters of the conditioning variable. Moving away from the normal distribution under the hidden truncation model will increase the difficulty in model fitting and related statistical inference. In search for a simpler non-normal model involving a two-sided hidden truncation, the bivariate exponential distribution proposed by Arnold and Strauss [12] is selected for illustrative purposes. As the statistical parameter estimation of a hidden truncation model is challenging, especially when there is a lack of information on the truncation limits in the

model, we aim to provide a feasible approach based on the Bayesian paradigm that can be used in practical situations.

The rest of the paper is organized as follows. In Section 2, the mathematical derivation for a basic two-component model of hidden truncated distributions and the mathematical derivation of a two-sided hidden truncated bivariate exponential (HTBEXP, in short) model are provided. Section 3 provides the details of the Bayesian inference adopted in this paper for the HTBEXP model developed in Section 2 under both an informative and non-informative and improper priors set-up. Section 4 explores the Bayes factor to deal with model selection procedures out of all possible candidate models. In Section 5, a Monte Carlo simulation study under the Bayesian paradigm is used to evaluate the performance of the proposed Bayesian estimation method. For illustrative purposes, a real-life dataset is re-analyzed in Section 6 to exhibit the efficacy of the proposed estimation strategy. Finally, some concluding remarks are presented in Section 7.

2. Hidden Truncation in Arnold–Strauss Bivariate Exponential Model

Let (X, Y) be a two-dimensional absolutely continuous random vector. Consider the conditional distribution of X given $Y \in M$, where M is a Borel set in $\mathbb{R}$. In a hidden truncation model, variable X is the variable of interest, and variable Y is a related unobservable variable that may be related to X. Let $f_{X,Y}(x, y)$ be the joint probability density function (PDF) of the random vector (X, Y), and let $f_X(x)$ and $f_Y(y)$ be the marginal PDFs of random variables X and Y, respectively. The conditional PDF of X given $Y \in M$ can be expressed as

$$f_{X|Y\in M}(x) = f_X(x)\left[\frac{\Pr(Y \in M|X = x)}{\Pr(Y \in M)}\right]. \tag{1}$$

The following three forms of hidden truncation models can be considered:

(i) Truncation from below (also known as lower truncation): $M = (d_1, \infty)$, where $-\infty < d_1 < \infty$ is the lower truncation point;
(ii) Truncation from above (also known as upper truncation): $M = (-\infty, d_2)$, where $-\infty < d_2 < \infty$ is the upper truncation point;
(iii) Two-sided truncation: $M = (d_1, d_2]$, where d_1 and d_2 are the lower and upper truncation points, respectively, with $-\infty < d_1 < d_2 < \infty$.

For a two-sided hidden truncation, observations are only available for X's whose corresponding concomitant variable Y is in $(d_1, d_2]$, for $-\infty < d_1 < d_2 < \infty$. Hence, Equation (1) reduces to

$$f_{X|Y\in(d_1,d_2]}(x|d_1 < Y \le d_2) = f_X(x)\left[\frac{\Pr(d_1 < Y \le d_2|X = x)}{\Pr(d_1 < Y \le d_2)}\right]. \tag{2}$$

This type of hidden truncation model is characterized as follows:

- The underlying marginal distribution of X can be specified by the PDF $f_X(x)$;
- The conditional distribution of Y given $X = x$ can be specified by the conditional PDF $f_{Y|X}(y|x)$;
- The specified values of truncation points are denoted by d_1 and d_2;
- There could be some other model parameters in addition to d_1 and d_2.

For the conditional PDF in Equation (2), if we consider $d_1 \to -\infty$ and $d_2 \to +\infty$, the PDF reduces to the unconditional marginal PDF of X. The shape of the conditional PDF will be more sensitive for smaller values of the lower truncation point d_1 compared to small or large values of the upper truncation point d_2. From the outset, the hidden truncation from below will not augment the original model since the resulting density will again be a member of the same family of distributions with only a reparametrization of the parent model. Consequently, we focus primarily on the hidden truncation from both sides in the more general framework, as the hidden truncation from below and/or from above will be a particular case of two-sided hidden truncation.

In this paper, for the applications in situations with random variables having non-negative supports (e.g., the survival and reliability analyses), a two-dimensional random vector (X, Y) with non-negative coordinates that follows the Arnold–Strauss bivariate exponential (ASBE) distribution is considered. The joint PDF of the random vector (X, Y) of ASBE distribution is

$$f_{X,Y}(x, y) = K \exp[-(ax + by + cxy)], \text{ for } x > 0,\ y > 0,\ a > 0,\ b > 0,\ c > 0, \tag{3}$$

where K is the normalizing constant defined as

$$K = \left[-c \exp\left(\frac{ab}{c}\right) Ei\left(-\frac{ab}{c}\right)\right]^{-1}. \tag{4}$$

Here, $Ei(x)$ is the exponential integral function defined by $Ei(x) = \int_{-\infty}^{x} t^{-1} \exp(t) dt$. The associated marginal PDF of X will be

$$f_X(x) = \frac{K \exp(-ax)}{b + cx}, \quad x > 0, \tag{5}$$

and the marginal density of Y will be

$$f_Y(y) = \frac{K \exp(-by)}{a + cy}, \quad y > 0.$$

Therefore, the normalizing constant K can also be defined as

$$K = \left[\int_0^\infty \frac{\exp(-ax)}{b + cx} dx\right]^{-1} = \left[\int_0^\infty \frac{\exp(-by)}{a + cy} dy\right]^{-1}.$$

Since $\Pr(Y > y | X = x)$ for each fixed $X = x$ is monotonically decreasing in x for all possible values of y, the joint density in Equation (3) is negative quadrant dependent, and hence it has a negative correlation coefficient. In the context of hidden truncation, we assume that the random variable of interest X and the hidden covariable Y are dependent non-negative random variables that follow the ASBE distribution.

Note that the conditional PDF of Y given $X = x$ is

$$f_{Y|X=x}(y | X = x) = (b + cx) \exp[-(b + cx)y], \ \ y > 0.$$

Next, a two-sided hidden truncation for the random variable Y, say $d_1 < Y \leq d_2$, where $0 \leq d_1 < d_2 < \infty$ is considered. We can obtain

$$\begin{aligned} \Pr(d_1 < Y \leq d_2 | X = x) &= \int_{d_1}^{d_2} (b + cx) \exp[-(b + cx)y] dy \\ &= \exp[-(b + cx)d_1] - \exp[-(b + cx)d_2], \end{aligned}$$

and

$$\Pr(d_1 < Y \leq d_2) = \int_{d_1}^{d_2} \frac{K \exp(-by)}{a + cy} dy. \tag{6}$$

Therefore, the conditional PDF of X given $d_1 < Y \leq d_2$, can be expressed as

$$
\begin{aligned}
f(x|d_1 < Y \leq d_2) &= f_X(x)\left[\frac{f_{Y|X=x}(y|X=x)}{\Pr(d_1 < Y \leq d_2)}\right] \\
&= \frac{\exp(-ax)\{\exp[-(b+cx)d_1] - \exp[-(b+cx)d_2]\}}{(b+cx)\int_{d_1}^{d_2}\frac{\exp(-by)}{a+cy}dy} \\
&= \frac{\exp(-ax)\{\exp[-(b+cx)d_1] - \exp[-(b+cx)d_2]\}}{(b+cx)M(a,b,c,d_1,d_2)}, \quad x > 0, \quad (7)
\end{aligned}
$$

where

$$
M(a,b,c,d_1,d_2) = \int_{d_1}^{d_2}\frac{\exp(-by)}{a+cy}dy \quad (8)
$$

is a constant depending only on the parameters and not the random variable X. Figure 1 presents the hidden truncated PDFs of X given $d_1 < Y \leq d_2$ for different values of the parameters a, b, c, d_1, and d_2 based on the ASBE distribution along with the non-truncated exponential distribution.

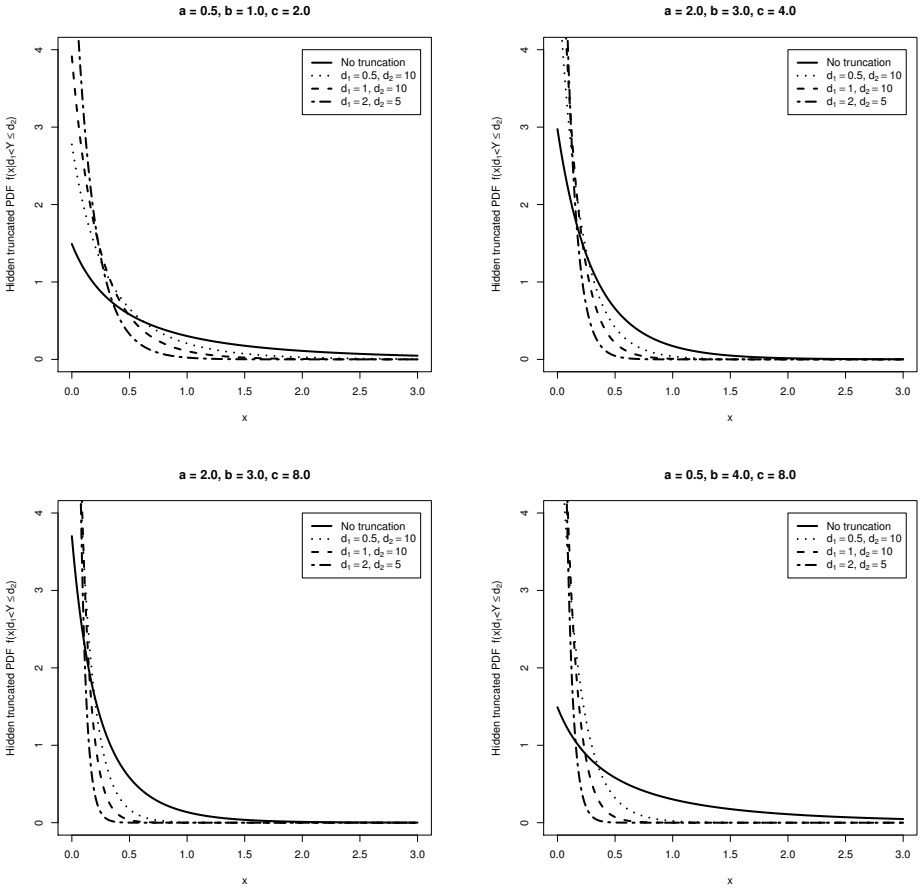

Figure 1. Hidden truncated PDFs of X given $d_1 < Y \leq d_2$ for different values of d_1 and d_2 based on the ASBE distribution along with the non-truncated exponential distribution.

From Figure 1, the hidden truncated PDFs are always right-skewed with different intensities for given choices of a, b, and c along with the truncated parameters $0 \leq d_1 < d_2 < \infty$. In Section 3, we discuss in detail the Bayesian inference of the density given in Equation (7).

3. Bayesian Inference

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from the hidden-truncated distribution with the PDF in Equation (7), and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ be the observed values of $\boldsymbol{X}$. Based on $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the associated likelihood function can be expressed as

$$\begin{aligned} & L(a, b, c, d_1, d_2|\boldsymbol{x}) \\ = & \frac{1}{[M(a, b, c, d_1, d_2)]^n} \prod_{i=1}^{n} \frac{\exp(-ax_i)\{\exp[-(b + cx_i)d_1] - \exp[-(b + cx_i)d_2]\}}{b + cx_i}. \end{aligned} \quad (9)$$

This section discusses the choices for the priors of the model parameters. Due to the condition that $d_1 < d_2$, we consider dependent priors for parameters d_1 and d_2. For parameters a, b, and c, the following two scenarios are considered:

- Independent priors for parameters a, b, and c;
- Dependent priors for parameters a, b, and c.

3.1. Prior Specifications

3.1.1. Independent Priors for the Parameters a, b, and c

When the underlying distribution follows an exponential distribution, it is common that a gamma-type distribution is used as a prior distribution in a subjective Bayesian framework. Since the ASBE distribution is a member of the exponential family, the conjugacy of priors is expected. Consider a random variable that follows a gamma distribution with shape parameter α and rate parameter β, denoted as $Gamma(\alpha, \beta)$, with the following PDF

$$g(w; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} \exp(-\beta w),\ w > 0,\ \alpha > 0,\ \beta > 0.$$

The following independent priors for parameters a, b, c, and dependent priors for d_1, and d_2 are considered:

- The prior for parameter a

$$\pi_1(a) \sim Gamma(\alpha_1, \beta_1).$$

- The prior for parameter b

$$\pi_2(b) \sim Gamma(\alpha_2, \beta_2).$$

- The prior for parameter c

$$\pi_3(c) \sim Gamma(\alpha_3, \beta_3).$$

- For the upper and lower truncation points d_1 and d_2, since $d_1 < d_2$, the following dependent priors are considered:

$$\begin{aligned} \pi_4(d_1) & \propto \frac{1}{(1 + d_1)^2} I(0 < d_1 < d_2), \\ \pi_5(d_2|d_1) & = \frac{1}{\delta - d_1} I(d_1 < d_2 < \delta). \end{aligned}$$

Remark 1. *Regarding the prior specifications on d_1 and d_2, the marginal prior distribution for d_1 is assumed to be a truncated distribution with support $(0, d_2]$. A uniform distribution for $d_2|d_1$ on the interval $[d_1, \delta]$ is used, where δ is a hyperparameter that should be chosen properly for the underlying Bayesian analysis. For instance, the informed expert can provide a judicious choice of hyperparameters. These conditional priors for the two truncation points may not be the optimal choice of priors, but among all the other different priors that we have considered, this one seems to perform satisfactorily.*

Based on the likelihood function in Equation (9), along with the joint prior distribution under *independent a priori*, and with the dependent prior choices for d_1 and d_2, the joint posterior distribution of the five model parameters can be expressed as

$$\begin{aligned} & h(a,b,c,d_1,d_2|\boldsymbol{x}) \\ \propto\ & L(a,b,c,d_1,d_2|\boldsymbol{x}) \times \pi_1(a) \times \pi_2(b) \times \pi_4(c) \times \pi_4(d_1) \times \pi_5(d_2|d_1) \\ & \frac{1}{[M(a,b,c,d_1,d_2)]^n} \prod_{i=1}^{n} \frac{\exp(-ax_i)\{\exp[-(b+cx_i)d_1] - \exp[-(b+cx_i)d_2]\}}{b+cx_i} \\ \times\ & g(a;\alpha_1,\beta_1)g(b;\alpha_2,\beta_2)g(c;\alpha_3,\beta_3)\left[\frac{1}{(1+d_1)^2}\right]\left[\frac{1}{\delta-d_1}\right]. \end{aligned} \tag{10}$$

Note that the marginal posterior distribution of each of the model parameters can be derived from the joint posterior in Equation (10). For instance, the posterior distribution of a can be obtained as

$$p_1(a|\boldsymbol{x}) = \frac{1}{D}\int_0^{d_2}\int_{d_1}^{\delta}\int_0^{\infty}\int_0^{\infty} h(a,b,c,d_1,d_2|\boldsymbol{x})db\ dc\ dd_2\ dd_1,$$

where D is the normalizing constant defined as

$$D = \int_0^{d_2}\int_{d_1}^{\delta}\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} h(a,b,c,d_1,d_2|\boldsymbol{x})da\ db\ dc\ dd_2\ dd_1.$$

Similarly, the posterior distributions for the remaining other parameters can be obtained. It is worth pointing out that each of the marginal distributions involves the quantity $M(a,b,c,d_1,d_2)$ defined in Equation (8), which is a constant. Due to the presence of the term D, an accurate approximation of the expectation in calculating the posterior mean is needed, and a five-dimensional numerical integration will be involved here. In this work, Gibbs sampling is used to find the Bayes estimates (posterior means) of the model parameters. Detailed procedures associated with the Metropolis–Hastings (MH) algorithms will be presented in Section 3.2.

3.1.2. Dependent Priors for Parameters a, b and c

In reality, the model parameters may be dependent. Hence, it is more reasonable to consider dependent priors. Here, some general prior choices for the model parameters are considered. Suppose we have specific information (in the form of a prior belief) about any of the parameters. Then, one can build informative priors for the rest of the parameter(s). For instance, if we know that parameter a can take values between 2 and 4, then a reasonable prior distribution for the parameter a would be any priors with support on $(2,4)$, e.g., a uniform distribution on the interval $(2,4)$. Although the use of dependent priors will increase the complexity of our analysis, one may still want to consider such priors. For analytically intractable models as given in Equation (7), the corresponding posterior analysis can be efficiently performed by Markov Chain Monte Carlo (MCMC) algorithms. However, in a real-life scenario, it is common that one might not have explicit knowledge of the priors for each parameter and the dependency between these parameters. Specifically, this occurs when we consider a situation in which the prior belief on either of the parameters a, b, or c, or some of the parameters a, b, and c depend on the two unknown truncation parameters d_1 and d_2.

Instead of independent priors, dependent priors for parameters a, b, and c can be considered. In this scenario, we assume that the prior distributions of a, b, and c follow conditional gamma distributions. For example,

$$\begin{cases} \pi_1^*(a|b,c) \sim Gamma(\xi_1(b,c),\lambda_1(b,c)), \\ \pi_2^*(b|a,c) \sim Gamma(\xi_2(a,c),\lambda_2(a,c)), \\ \pi_3^*(c|a,b) \sim Gamma(\xi_3(a,b),\lambda_3(a,b)). \end{cases}$$

On the other hand, the same prior distributions for d_1 and d_2 that were defined in Section 3.1.1 are used.

Next, the joint prior PDF of parameters a, b, and c can be expressed as (see [13] for details)

$$\begin{aligned} f(a,b,c) \;=\; & \exp\{m_{000} - m_{100}a + m_{200}\ln a - m_{010}b + m_{110}ab - m_{210}b\ln a \\ & + m_{020}\ln b - m_{120}a\ln b + m_{220}\ln a\ln b - m_{001}c + m_{101}ca - m_{201}c\ln a \\ & + m_{011}cb - m_{111}abc - m_{021}c\ln b + m_{121}ca\ln b + m_{211}cb\ln a \\ & - m_{221}\ln b + m_{002}\ln c - m_{102}a\ln c + m_{202}\ln a\ln c - m_{012}b\ln c \\ & + m_{112}ab\ln c - m_{212}b\ln a\ln c - m_{022}\ln b\ln c - m_{122}a\ln b\ln c \\ & + m_{222}\ln a\ln b\ln c\}I(a>0)I(b>0)I(c>0). \end{aligned} \tag{11}$$

The family of PDFs in Equation (11) involves 27 hyperparametrs. A simplified sub-model, which still retains considerable flexibility, can be obtained by setting $m_{ijk} = 0$, whenever $i + j + k > 2$. Then, Equation (11) reduces to

$$\begin{aligned} f^*(a,b,c) \;=\; & \exp(m_{000} - m_{100}a + m_{200}\ln a - m_{010}b + m_{110}ab + m_{020}\ln b - m_{001}c \\ & + m_{101}ca + m_{011}cb + m_{002}\ln c)I(a>0)I(b>0)I(c>0). \end{aligned} \tag{12}$$

When the simplified 10-parameter model in Equation (12) is used, the joint posterior distribution from Equation (9) reduces to

$$h^*(a,b,c,d_1,d_2|\boldsymbol{x}) \propto L(a,b,c,d_1,d_2|\boldsymbol{x}) \times f^*(a,b,c) \times \pi_4(d_1) \times \pi_5(d_2|d_1). \tag{13}$$

3.2. Computation Algorithms

In this subsection, we present the detailed procedures for the MCMC algorithms in which the MH algorithms within Gibbs chains are utilized. Furthermore, different types of MH algorithms are adopted depending on the types of full conditional distributions. Since the procedures regarding the joint posterior distribution for dependent priors are similar, we present the posterior analysis based on only independent priors for brevity. From the joint posterior distribution in Equation (10), we can construct the full conditional distributions to proceed with utilizing the Gibbs sampling. The full conditional distributions are given as follows:

$$\begin{aligned} p(a|b,c,d_1,d_2,\boldsymbol{x}) \;\propto\; & \frac{1}{[M(a,b,c,d_1,d_2)]^n}a^{\alpha_1-1}\exp\left\{-\left[\sum_{i=1}^{n}x_i+\beta_1\right]a\right\}I(a>0); \\ p(b|a,c,d_1,d_2,\boldsymbol{x}) \;\propto\; & \frac{1}{[M(a,b,c,d_1,d_2)]^n}\prod_{i=1}^{n}\left[\frac{\exp\{-(b+cx_i)d_1\}-\exp\{-(b+cx_i)d_2\}}{b+cx_i}\right] \\ & \times b^{\alpha_2-1}\exp\{-\beta_2 b\}I(b>0); \\ p(c|a,b,d_1,d_2,\boldsymbol{x}) \;\propto\; & \frac{1}{[M(a,b,c,d_1,d_2)]^n}\prod_{i=1}^{n}\left[\frac{\exp\{-(b+cx_i)d_1\}-\exp\{-(b+cx_i)d_2\}}{b+cx_i}\right] \\ & \times c^{\alpha_3-1}\exp\{-\beta_3 c\}I(c>0); \\ p(d_1|a,b,c,d_2,\boldsymbol{x}) \;\propto\; & \frac{1}{[M(a,b,c,d_1,d_2)]^n}\prod_{i=1}^{n}\left[\frac{\exp\{-(b+cx_i)d_1\}-\exp\{-(b+cx_i)d_2\}}{b+cx_i}\right] \\ & \times\frac{1}{(1+d_1)^2}I(0<d_1<d_2); \\ p(d_2|a,b,c,d_1,\boldsymbol{x}) \;\propto\; & \frac{1}{[M(a,b,c,d_1,d_2)]^n}\prod_{i=1}^{n}\left[\frac{\exp\{-(b+cx_i)d_1\}-\exp\{-(b+cx_i)d_2\}}{b+cx_i}\right] \\ & \times\frac{1}{\delta-d_1}I(d_1<d_2<\delta). \end{aligned} \tag{14}$$

Note that none of the full conditional distributions has a closed-form expression. Hence, we propose using MH algorithms within each Gibbs chain. For parameters a, b, and

c, a truncated normal distribution is used as a proposal density with a left-truncation point of zero. In particular, an adaptive MH algorithm that updates the variance of proposal distributions throughout the entire sampling process is used. In general, the proposal density of parameter θ ($\theta = a, b, c, d_1$, or d_2) is used as a (truncated) normal distribution, with the current value being its mean and the variance depending on forthcoming iterations. More specifically, suppose we want to update parameter θ in the $(h+1)^{th}$ step based on an initial value of $\theta^{(0)}$ and observed values of the first h iterations are denoted by $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(h)}$. The candidate value θ^* can be generated from a normal distribution with mean $\theta^{(h)}$ and variance $V^{(h)}$, where

$$V^{(h)} = \begin{cases} V^{(0)}, & \text{if } h = 0, \\ \gamma[\text{Var}\left(\theta^{(0)}, \ldots, \theta^{(h-1)}\right) + \zeta], & \text{if } h > 0. \end{cases} \tag{15}$$

Here, $V^{(0)}$ is an initial (could be arbitrary) variance of the proposal distribution of parameter θ, γ is the adjusting coefficient, and ζ is a small constant to prevent the variance in Equation (15) from shrinking to zero. One of the important aspects of the adaptive MH algorithm is that it updates its variance so that the parameter being sampled reaches a desired acceptance rate. It is known that the optimal acceptance rate for one parameter is 0.44, so we set $\gamma = 2.4/\sqrt{k}$, where k is the dimension of the parameter space. This ensures that the desired acceptance rate stays around 0.44; see [14] for pertinent details.

On the other hand, an independent MH algorithm is used to generate random variates from the full conditional distribution of d_1. The algorithm for generating a random variate from $p(d_1|\text{others})$ is described below. When $d_1^{(j)}$ is the current value of the Markov chain, updating $d_1^{(j+1)}$ is required. The steps of the algorithm are described as follows.

A1. Let q be the candidate density with a form:

$$q(d_1) = \frac{1+d_2}{d_2} \frac{1}{(1+d_1)^2} \text{ for } 0 < d_1 < d_2, \tag{16}$$

where $d_2 \equiv d_2^{(j)}$ is the value of the current state.

A2. Generate d_1^* from the PDF in Equation (16). Following Devroye [15], and the inverse cumulative distribution function (CDF) technique, d_1^* can be generated from the following equation:

$$d_1^* = \frac{1+d_2}{1+d_2(1-u)} - 1, \tag{17}$$

where u is a random variate from Unif$(0, 1)$.

A3. Next, the acceptance probability ω is calculated:

$$\omega(d_1^{(j)}, d_1^*) = \frac{p(d_1^*)q(d_1^{(j)})}{p(d_1^{(j)})q(d_1^*)},$$

where $p(\cdot)$ is the full conditional distribution of d_1 in Equation (14). Now, set $d_1^{(j+1)} = d_1^*$ if $u \leq \omega(d_1^{(j)}, d_1^*)$, and $d_1^{(j+1)} = d_1^{(j)}$ otherwise.

Finally, we employ the random-walk MH for updating d_2. When $d_2^{(j)}$ is given, we wish to update $d_2^{(j+1)}$. Generate d_2^* from Unif$(d_2^{(j)} - \epsilon, d_2^{(j)} + \epsilon)$ for $\epsilon > 0$. Next, we calculate the acceptance probability:

$$\omega(d_2^{(j)}, d_2^*) = \frac{p(d_2^*)I(d_1 < d_2^* < \delta)}{p(d_2^{(j)})I(d_1 < d_2^{(j)} < \delta)},$$

where $d_1 \equiv d_1^{(j)}$ is the value of the current state. If $u \leq \omega(d_2^{(j)}, d_2^*)$ and $I(d_1 < d_2^* < \delta)$, then set $d_2^{(j+1)} = d_2^*$. Otherwise, $d_2^{(j+1)} = d_2^{(j)}$, and $u \sim \text{Unif}(0,1)$.

Remark 2. *The efficacy of the MCMC algorithm involving analytically intractable models, such as the case here with the constant $M(a,b,c,d_1,d_2)$, can be a concern. For instance, the constant $M(a,b,c,d_1,d_2)$ is in an integral form, and the exact value may not be able to be obtained analytically. However, we can approximate $M(a,b,c,d_1,d_2)$ accurately for given values of a, b, c, d_1, and d_2 and perform the Bayesian computation with the MCMC strategy without any difficulty. Noticeably, the results based on the simulation study are supportive of this assertion. One may need to consider non-standard MCMC techniques for doubly intractable posteriors; see, for example, [16]. In the Section 4, we discuss the selection strategy to select among all possible candidate hidden truncation models using the Bayes factor, the most appropriate and plausible model.*

4. Selection Between Different Hidden Truncation Models

In this section, we consider the following four models for the hidden truncated bivariate exponential distribution:

- M_1: No hidden truncated model, i.e., $d_1 = 0$ and $d_2 = \infty$.
- M_2: A truncated from below model, i.e., $0 < d_1 < d_2 = \infty$.
- M_3: A truncate from above model, i.e., $0 = d_1 < d_2 < \infty$.
- M_4: A truncate from both sides model, i.e., $d_1 > 0$ and $d_2 < \infty$.

Let Θ_j denote the parameter space for model M_j $(j = 1,2,3,4)$. Hence, we have $\Theta_1 = (a,b,c)$, $\Theta_2 = (a,b,c,d_1)$, $\Theta_3 = (a,b,c,d_2)$, and $\Theta_4 = (a,b,c,d_1,d_2)$. Based on the observed values $\boldsymbol{x} = (x_1, x_2, \ldots x_n)$ from each of models M_1, M_2, and M_3, the likelihood functions under these three models are given respectively by

$$\begin{aligned}
L_1(a,b,c|\boldsymbol{x}) &= \frac{1}{[I_1(a,b,c)]^n} \prod_{i=1}^{n} \frac{\exp\{-ax_i\}}{b+cx_i}, \\
L_2(a,b,c,d_1|\boldsymbol{x}) &= \frac{1}{[I_2(a,b,c,d_1)]^n} \prod_{i=1}^{n} \frac{\exp\{-ax_i\}\exp\{-(b+cx_i)d_1\}}{b+cx_i}, \\
L_3(a,b,c,d_2|\boldsymbol{x}) &= \frac{1}{[I_3(a,b,c,d_2)]^n} \prod_{i=1}^{n} \frac{\exp\{-ax_i\}[1-\exp\{-(b+cx_i)d_2\}]}{b+cx_i}, \qquad (18)
\end{aligned}$$

where $I_1(a,b,c)$ is the same as $1/K$ in Equation (4) of the current text, and

$$I_2(a,b,c,d_1) = \int_{d_1}^{\infty} \frac{\exp\{-by\}}{a+cy} dy \text{ and } I_3(a,b,c,d_2) = \int_0^{d_2} \frac{\exp\{-by\}}{a+cy} dy.$$

For model M_4, the likelihood function $L(a,b,c,d_1,d_2|\boldsymbol{x})$ is presented in Equation (9).

Remark 3. *Regarding prior specifications for models M_1, M_2, and M_3, we use the same prior distributions utilized in the full model M_4 by deleting irrelevant parameters to estimate the corresponding parameters. Subsequently, we could obtain MCMC outputs in the three models that can also approximate marginal distributions.*

To choose the most plausible model based on the observed data, we consider using the Bayesian approach based on the Bayes factor. Suppose that there are q different models, namely, $M_1, \ldots, M_q$, any of which could be meaningful, and they contend with each other for model selection. If model M_i holds true, the data $\boldsymbol{x} = (x_1, \ldots, x_n)$ follow a parametric distribution with PDF $f_i(x|\boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is an unknown parameter vector for M_i. Let Θ_i be the parameter space for the parameter vector $\boldsymbol{\theta}_i$, where Θ_i may or may not be nested. Bayesian model selection proceeds by choosing a prior PDF $\pi_i(\boldsymbol{\theta}_i)$ for $\boldsymbol{\theta}_i$ under M_i, and a

prior model probability $p(M_i)$ of M_i being true for $i = 1, \ldots, q$. The posterior model of the probability that M_i is true can be obtained as (for pertinent details, see, [17])

$$P(M_i|\boldsymbol{x}) = \left[\sum_{j=1}^{q} \frac{p(M_j)}{p(M_i)} B_{ji}(\boldsymbol{x})\right]^{-1}, \tag{19}$$

where the Bayes factor $B_{ji}(\boldsymbol{x})$ of model M_j to M_i is defined by

$$B_{ji}(\boldsymbol{x}) = \frac{m_j(\boldsymbol{x})}{m_i(\boldsymbol{x})} = \frac{\int_{\Theta_j} f_j(\boldsymbol{x}|\theta_j)\pi_j(\theta_j)d\theta_j}{\int_{\Theta_i} f_i(\boldsymbol{x}|\theta_i)\pi_i(\theta_i)d\theta_i}, \tag{20}$$

where $m_i(\boldsymbol{x})$ is called the marginal or predictive density of $\boldsymbol{x}$ under M_i. Subsequently, the model with the largest posterior probability in Equation (19) becomes the most plausible model. It is customary to use vague model probabilities, i.e., $p(M_i) = 1/q$ for $i = 1, \ldots, q$. Then, Equation (19) becomes $P(M_i|\boldsymbol{x}) = [\sum_{j=1}^{q} B_{ji}]^{-1}$. Clearly,

$$B_{ii} = 1 \text{ and } B_{ij} = B_{ji}^{-1}, \text{ for } i, j = 1, \ldots, q. \tag{21}$$

Meanwhile, we use the method proposed by Newton and Raftery [18] to evaluate the marginals of the four models. The procedure is described as follows. Observe that the posterior distribution can be represented as $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto L_j(\boldsymbol{\theta}|\boldsymbol{x}) \times \pi_j(\boldsymbol{\theta})$ under model M_j for $j = 1, 2, 3, 4$. For notational simplicity, let $\boldsymbol{\theta} = (a, b, c, d_1, d_2)$ under the full model M_4. Now, let $\{\boldsymbol{\theta}^{(g)}\} \equiv \{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(G)}\}$ be G, drawing from the posterior density obtained by conventional MCMC methods such as the Gibbs sampler. Newton and Raftery [18] suggested that the marginal can be estimated as

$$\hat{m}_4(\boldsymbol{x}) = \left\{\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{L_4(\boldsymbol{x}|\boldsymbol{\theta}^{(g)})}\right)\right\}^{-1}. \tag{22}$$

Note that the estimate in Equation (22) is the harmonic mean of the likelihood evaluated at MCMC samples.

5. Monte Carlo Simulation Studies

In this section, Monte Carlo simulation studies are used to evaluate the Bayesian parameter estimation and model selection procedures proposed in this paper for the hidden truncation models. First, a Monte Carlo simulation study is used to evaluate the performance of Bayes estimates for the model parameters of the probability distribution given in Equation (7). For illustrative purposes, we consider the true values of the model parameters to be $(a, b, c, d_1, d_2) = (2, 3, 4, 1, 10)$, and three different sample sizes n = 100, 200, and 300. For each simulated dataset, the posterior means, posterior medians, and the associated 95% highest posterior density (HPD) credible intervals of the model parameters are computed. After collecting the posterior estimates based on a total of 200 replications, the performance of the point estimates is evaluated based on the average posterior mean, coverage probability (CP), and average width (AW) of the 95%. The total and burn-in MCMC iterations of the Bayesian estimation procedure are set to be 20,000 and 5000, respectively. Regarding the hyperparameters of the prior distributions on a, b, and c, we set $(\alpha_1, \beta_1) = (4, 2)$, $(\alpha_2, \beta_2) = (9, 3)$, and $(\alpha_3, \beta_3) = (16, 4)$, in order to obtain the mean of each prior distribution to be the same as the true value and the variances as one for the three prior gamma distributions. For the hyperparameter(s) of the distribution on $d_2|d_1$, we set $\delta = 20$. Moreover, regarding the initial variances in Equation (15) of the zero-truncated normal distributions, we set a standard deviation of two for each of a, b, and c. Table 1 provides the average posterior means, the mean square errors (MSEs) of the posterior means, the average posterior variances, the CPs, and AWs for the model parameters based on 200 simulations.

Table 1. Simulation results when the data are generated from M_4 with parameter values $(a, b, c, d_1, d_2) = (2, 3, 4, 1, 10)$.

Sample Size	Posterior Estimate	Parameters				
		a	b	c	d_1	d_2
100	Average posterior mean	2.443	2.952	4.102	0.903	10.504
	MSE of posterior mean	0.203	0.011	0.017	0.054	0.279
	Average posterior variance	1.411	0.974	0.938	0.192	30.125
	Coverage probability (CP)	100%	100%	100%	100%	100%
	Average width (AW)	4.348	3.715	3.695	1.621	17.968
200	Average posterior mean	2.458	2.951	4.092	0.898	10.495
	MSE of posterior mean	0.217	0.021	0.023	0.031	0.264
	Average posterior variance	1.442	0.962	0.921	0.175	30.159
	Coverage probability	100%	100%	100%	100%	100%
	Average width	4.411	3.682	3.665	1.585	19.955
300	Average posterior mean	2.433	2.960	4.098	0.901	10.495
	MSE of posterior mean	0.210	0.028	0.031	0.027	0.269
	Average posterior variance	1.407	0.957	0.901	0.164	30.125
	Coverage probability (CP)	100%	100%	100%	100%	100%
	Average width (AW)	4.371	3.674	3.610	1.552	17.944

From Table 1, we observe that all the CPs are 100%, which indicates that the standard errors of the estimates are relatively large. The posterior means of b, c, d_1, and d_2 are close to the true values for decently large sample sizes. However, the posterior means for a are overestimated with large average widths for the HPD intervals. We also observe that there are no dramatic improvements in parameter estimation accuracy when the sample size increases. Since we are dealing with subjective prior choices, there are infinitely many combinations on the hyperparameters, and the performance of the Bayesian estimation procedure depends on the choice of the hyperparameters. For instance, based on our preliminary study (results are not shown here), the performance can be poor when prior distributions with large prior variances are used. Although the choices of hyperparameter values used in this simulation study may not be optimal, these choices of hyperparameter values provide satisfactory results among the choices we examined.

In the second Monte Carlo simulation study, the performance of the model selection procedure is evaluated when the four models M_1–M_4 are considered candidate models. In this simulation study, data are generated from the four models M_1–M_4, and the model selection method based on Bayes factors presented in Section 4 is applied to each simulated dataset. Then, we obtain the proportions to determine which model has the largest posterior model probability. Recall that the posterior probabilities can be computed by Equation (19) in conjunction with Equation (22). For comparative purposes, we compute the average posterior probabilities in two ways by assuming the equal prior model probability of $1/4$ for each model M_1–M_4:

I. The average of posterior probabilities, no matter which model is selected out of 500 replications (denoted as Post. Prob. I in Table 2);
II. The average of posterior probabilities is based only on the samples possessing the highest posterior probability in each replication (denoted as Post. Prob. I in Table 2). For example, the "Post. Prob. II" when the true model is M_1 in Table 2, 0.265 for contending model M_1 is computed based on 125 samples that model M_1 is selected.

Based on 500 simulated samples from each model, the above two types of posterior probabilities, the frequency and proportion of selecting a particular model, and the average posterior means (medians) based only on selected models for each model parameter are presented in Table 2. The numerical summaries in Table 2 with the same prior specifications and iterations are in accordance with the adopted MCMC procedures described in Section 3.2.

Table 2. Posterior probabilities and Bayes estimates for different models.

True Model	Posterior Probability & Parameter Estimate	Contending Models			
		M_1	M_2	M_3	M_4
M_1	Post. Prob. I	0.251	0.249	0.251	0.249
	Freq. (Prop.) of selection	125 (0.250)	142 (0.284)	96 (0.192)	137 (0.274)
	Post. Prob. II	0.265	0.261	0.265	0.261
	Post. mean (median) for a	1.912 (1.910)	1.339 (1.350)	1.971 (1.964)	1.372 (1.379)
	Post. mean (median) for b	2.777 (2.660)	3.320 (3.210)	2.771 (2.657)	3.344 (3.227)
	Post. mean (median) for c	4.158 (4.088)	3.471 (3.392)	4.179 (4.101)	3.480 (3.408)
	Post. mean (median) for d_1	$d_1 = 0$	0.250 (0.224)	$d_1 = 0$	0.251 (0.220)
	Post. mean (median) for d_2	$d_2 = \infty$	$d_2 = \infty$	10.000 (10.000)	10.000 (10.000)
M_2	Post. Prob. I	0.239	0.261	0.239	0.261
	Freq. (Prop.) of selection	46 (0.092)	186 (0.372)	44 (0.088)	224 (0.448)
	Post. Prob. II	0.271	0.270	0.278	0.270
	Post. mean (median) for a	5.301 (5.315)	2.033 (1.889)	5.299 (5.308)	2.043 (1.899)
	Post. mean (median) for b	2.377 (2.250)	3.060 (2.951)	2.338 (2.207)	3.062 (2.939)
	Post. mean (median) for c	4.460 (4.376)	3.789 (3.691)	4.483 (4.402)	3.786 (3.716)
	Post. mean (median) for d_1	$d_1 = 0$	1.127 (1.113)	$d_1 = 0$	1.131 (1.103)
	Post. mean (median) for d_2	$d_2 = \infty$	$d_2 = \infty$	10.000 (9.999)	10.000 (10.000)
M_3	Post. Prob. I	0.251	0.250	0.250	0.249
	Freq. (Prop.) of selection	105 (0.210)	140 (0.280)	109 (0.218)	146 (0.292)
	Post. Prob. II	0.264	0.260	0.265	0.261
	Post. mean (median) for a	1.941 (1.934)	1.358 (1.368)	1.934 (1.922)	1.360 (1.368)
	Post. mean (median) for b	2.759 (2.645)	3.321 (3.206)	2.763 (2.643)	3.339 (3.227)
	Post. mean (median) for c	4.161 (4.087)	3.488 (3.416)	4.186 (4.109)	3.482 (3.406)
	Post. mean (median) for d_1	$d_1 = 0$	0.250 (0.222)	$d_1 = 0$	0.247 (0.217)
	Post. mean (median) for d_2	$d_2 = \infty$	$d_2 = \infty$	10.000 (9.999)	10.000 (10.000)
M_4	Post. Prob. I	0.239	0.261	0.239	0.261
	Freq. (Prop.) of selection	45 (0.090)	205 (0.410)	42 (0.084)	208 (0.416)
	Post. Prob. II	0.274	0.270	0.276	0.270
	Post. mean (median) for a	5.224 (5.230)	2.033 (1.889)	5.287 (5.229)	2.042 (1.897)
	Post. mean (median) for b	2.348 (2.220)	3.055 (2.945)	2.342 (2.207)	3.059 (2.949)
	Post. mean (median) for c	4.482 (4.393)	3.787 (3.692)	4.481 (4.399)	3.790 (3.695)
	Post. mean (median) for d_1	$d_1 = 0$	1.139 (1.127)	$d_1 = 0$	1.122 (1.101)
	Post. mean (median) for d_2	$d_2 = \infty$	$d_2 = \infty$	10.000 (10.001)	10.000 (10.000)

From Table 2, we observe that the parameter estimates based on either the posterior mean or the posterior median of the four models yield decently reasonable results for all cases considered. To select the most plausible model, although the procedure based on the Bayes factor (posterior probability) does not capture the true model in all the cases, the procedure based on the Bayes factor tends to select model M_4, which is the most general hidden truncation model among the four models with the largest number of parameters.

Since the number of parameters models M_1–M_4 are different, instead of evaluating the performance based on the parameter estimates, we consider the estimation of the conditional mean $E(X|w_1 < Y < w_2)$, where w_1 and w_2 are pre-specified values based on the model and the values of d_1 and d_2 (e.g., for model M_1, $w_1 = 0$ and $w_2 = \infty$), which

is a function of the model parameters. Specifically, for given values of w_1 and w_2, the conditional expectation is

$$\begin{aligned}\tau(a,b,c) &= \int_0^\infty x f_{X|y}(x|w_1 < Y < w_2)dx \\ &= \int_0^\infty \frac{xe^{-ax}[\exp\{-(b+cx)w_1\} - \exp\{-(b+cx)w_2\}]}{(b+cx)M^*(a,b,c,w_1,w_2)},\end{aligned}$$

where

$$M^*(a,b,c,w_1,w_2) = \int_{w_1}^{w_2} \frac{e^{-by}}{a+cy}dy.$$

Let R be the total number of simulations (i.e., $R = 500$ in this study) and r_i be the number of samples that model M_i is selected with the largest posterior model probability ($i = 1,2,3,4$), where $R = r_1 + r_2 + r_3 + r_4$. Suppose $\hat{\tau}_{ij}(a,b,c)$ is the estimate of $\tau(a,b,c)$ based on the j-th sample, where model M_i is selected as the most plausible model for $i = 1,2,3,4$, $j = 1,2,\ldots,r_i$. The MSEs based on the simulated data corresponding to selected model M_i, denoted as MSE_i, can be calculated as

$$MSE_i = \frac{1}{r_i}\sum_{j=1}^{r_i}[\hat{\tau}_{ij}(a,b,c) - \tau(a,b,c)]^2. \tag{23}$$

To illustrate the effect on the performance of estimating the conditional expectation with the model selection procedure, the MSE with the model selection procedure is calculated as

$$MSE_{MS} = \frac{1}{R}\sum_{i=1}^{4}\sum_{j=1}^{r_1}[\hat{\tau}_{ij}(a,b,c) - \tau(a,b,c)]^2. \tag{24}$$

The simulated values of MSE_i ($i = 1,2,3,4$) and MSE_{MS} for estimating the conditional expectation associated with the model selection perspectives are presented in Table 3. Note that the proportions of selecting the correct models with the highest posterior probability are all the same as in Table 2. As expected, the value of MSE_i is the smallest if M_i is the true model in most cases. The simulation results in Table 3 show that the model selection procedure can reduce the risk of mis-specifying the underlying model. On the other hand, we observe that the values of MSE_1 and MSE_3 are similar, while the values of MSE_2 and MSE_4 are similar. Note that models M_1 and M_3 assume $d_1 = 0$, while M_2 and M_4 assume $d_1 > 0$ with an unknown value. This observation suggests that determining whether $d = 0$ may provide further information to improve the estimation of the conditional expectation.

Table 3. Mean square errors for estimating the conditional expectation.

True Model	**Contending Models**				**Model Selection MSE_{MS}**
	M_1	M_2	M_3	M_4	
	MSE_1	MSE_2	MSE_3	MSE_4	
M_1	0.00069	0.00085	0.00068	0.00086	0.00081
M_2	0.00016	0.00010	0.00016	0.00011	0.00012
M_3	0.00069	0.00085	0.00068	0.00085	0.00081
M_4	0.00016	0.00011	0.00015	0.00010	0.00012

6. Illustrative Example

In this section, the proposed methodologies in this paper are illustrated using a real dataset comprised of independently collected quadruplets of the redshift and the apparent magnitude of a quasar object previously analyzed by Efron and Petrosian [3]. The dataset

is available in the R package DTDA version 3.0.1 [19] named `Quasars`. The original data consist of $n = 210$ observations with the following variables $(z_i, m_i, d_{1i}, d_{2i})$ for $i = 1, \ldots, n$, where z_i denotes the redshift of the i-th quasar, m_i denotes its apparent magnitude, and the two values d_{1i} and d_{2i} indicate lower and upper truncation bounds corresponding to the apparent magnitude, respectively. The variable of interest is the transformed logarithm of luminosity values provided in the first column of the dataset `Quasars`, i.e., $V_i = t(z_i, m_i)$, where t is a transformation that depends on the cosmological model assumed (see [20,21] for a detailed description). The anti-logarithm transformation $X_i = \exp(V_i)$ is considered such that the support of the random variable X_i is $[0, \infty)$. We conjecture that the dataset has been subjected to a two-sided truncation and the PDF in Equation (7) would be appropriate to model this dataset, where the observation X_i is subjected to a two-sided truncation by other covariates, and the same truncation points apply for X_i, $i = 1, 2, \ldots, n$, i.e., $d_{1i} \equiv d_1$ and $d_{2i} \equiv d_2$.

Based on the model in Equation (7), and using the Bayesian methodologies proposed in this paper, the results presented in Table 4 are obtained. Regarding the hyperparameters of the prior distributions on a, b, and c, we set $(\alpha_j, \beta_j) = (4, 1)$ for $j = 1, 2, 3$ because we have no information on these parameters. The same value of $\delta = 20$ as in the Monte Carlo simulation studies is used. In Table 4, we report the posterior probabilities for each of the four models M_1–M_4, and the posterior means and medians (presented in the parenthesis) along with the corresponding 95% HPD intervals. From a model selection perspective, model M_2 is selected with the largest posterior probability of 0.305, assuming equal prior model probability for all four models. We notice that M_4 has the second largest posterior probability of 0.283, which agrees with our observation of the similarity of models M_2 and M_4 in the simulation results. Regarding the posterior inference, we observe that the estimates of a, b, and c have some patterns. The estimates under models M_1 and M_3 are close to each other, while those values are close to each other when models M_2 and M_4 are applied. Thus, the HPD intervals are changed accordingly. Figure 2 shows the trace plots for all parameters when the four models are adapted, respectively. Overall, the plots for a, b, and c are decently well behaved, while there are some fluctuations in the plots of d_1 and d_2.

Table 4. Posterior probabilities, posterior means (medians), and 95% HPD intervals for quasar data.

Model	**Posterior Probability**	**Parameters**				
		a	b	c	d_1	d_2
M_1	0.203	0.535 (0.536) (0.433, 0.644)	7.044 (6.741) (2.645, 11.889)	0.994 (0.893) (0.166, 2.067)	$d_1 = 0$ $d_1 = 0$	$d_2 = \infty$ $d_2 = \infty$
M_2	**0.305**	0.435 (0.450) (0.195, 0.614)	6.261 (6.000) (1.969, 10.649)	0.666 (0.588) (0.060, 1.388)	0.238 (0.170) (0, 0.686)	$d_2 = \infty$ $d_2 = \infty$
M_3	0.209	0.537 (0.536) (0.436, 0.646)	7.073 (6.741) (2.588, 11.989)	0.992 (0.889) (0.158, 2.100)	$d_1 = 0$ $d_1 = 0$	9.499(9.499) (9.020, 9.971)
M_4	0.283	0.439 (0.452) (0.226, 0.628)	6.500 (6.232) (2.392, 11.198)	0.722 (0.634) (0.117, 1.527)	0.210 (0.148) (0, 0.661)	9.500 (9.501) (9.006, 9.956)

We want to highlight the fact that the parameter estimates/values of parameter d_1 and d_2 (i.e., the last two columns) in Table 4 and for the Model M_2 having the maximum posterior probability confirm the fact that the data (alias the main study variable, X_i) have been subject to a one-sided upper truncation, as the value of $d_1 \neq 0$. In this data application, we have explored all possible scenarios of hidden truncation (including zero truncation) and then adopted strategy searches for the best-case scenario under the Bayesian paradigm.

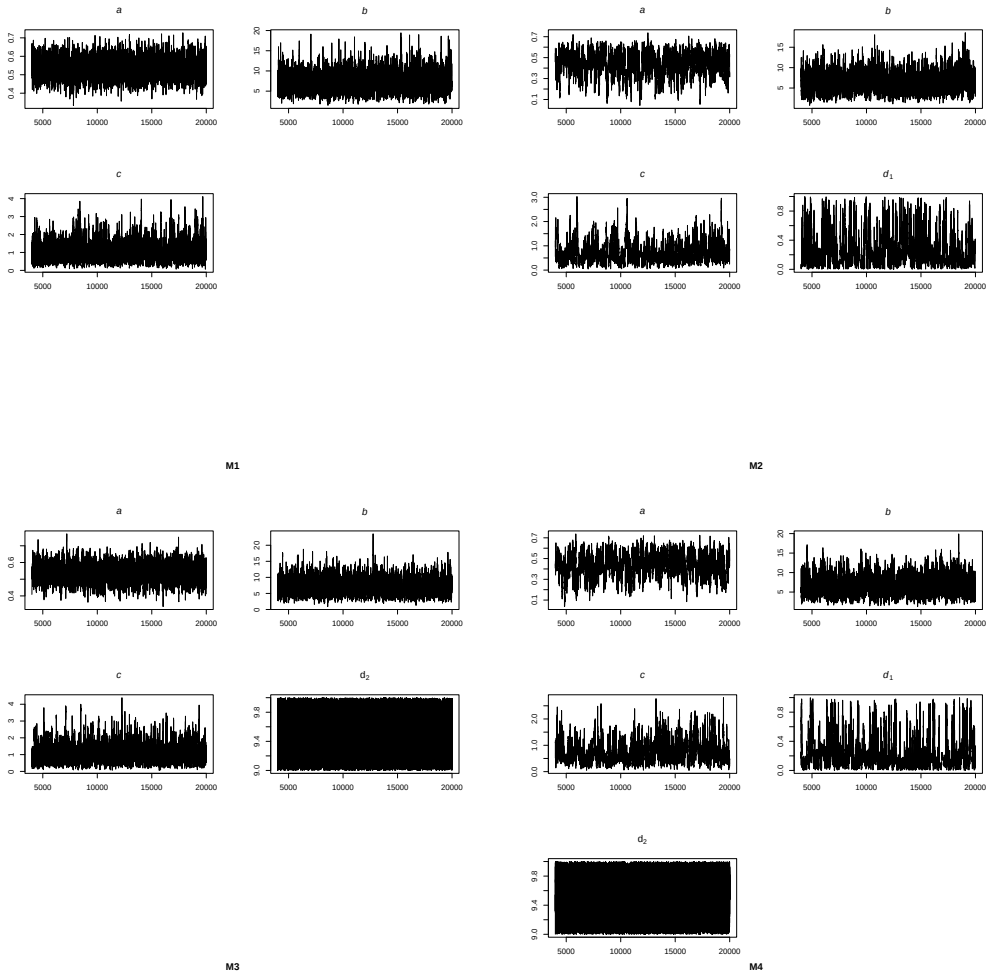

Figure 2. Traceplots of the estimates under different models.

7. Concluding Remarks

In this article, we explore the features of a two-sided hidden truncated model in terms of estimation under the Bayesian paradigm and, most importantly, detecting with a desired level of accuracy whether or not the data have been subject to hidden truncation starting with a simple model in two dimensions, namely, the bivariate exponential distribution proposed by Arnold and Strauss [12]. Indeed, there are several other versions of the classical bivariate exponential distribution, but they may have a singular part (i.e., if two random variables Y_1 and Y_2 follow a bivariate distribution, there is a positive probability that $Y_1 = Y_2$) [22] (see, for example, [23]), which makes it more difficult to consider from real-world and mathematical perspectives. For this reason, we resort to a simple model that is an absolutely continuous statistical distribution, develop the corresponding hidden-truncation model, and provide some feasible model-fitting methodologies that can be used in practice. Consequently, inferential aspects under the Bayesian paradigm for the parameters of a doubly (two-sided) hidden truncation model are still in their infancy stage and have not been explored in detail to the best of our knowledge. Statistical inference of the hidden truncation models with unknown truncation point(s) is a challenging problem, especially when there is very little information about the truncation point(s) in the observed sample. One can expect to mimic the inferential results obtained in this paper to other types of hidden truncated bivariate exponential models, albeit with computational complexity and possibly with a different set of informative and non-informative priors that might be challenging. Needless to say, the associated computational complexity and judicious selection of prior choices, especially for the truncation points, are the natural hindrance to developing the methodology and theory. Inference under the Bayesian framework, as

discussed in the simulation study as well as in the real-data application, is encouraging in the sense that the Bayesian estimates of the parameters are reasonably good under both settings. Efficient estimation for multiple constraints models, i.e., the estimation of model parameters for a hidden truncated model under a multi-component set-up, where, for example, the main study variable Y is observed only when the associated concomitant variables, say, X_j are truncated from below, such as $X_j < c_j$ (or $\ell_j < X_j < d_j$), will be of great interest in the context of several real-life applications. However, the real-life applicability of such models and their analytical traceability are the two key factors that should be addressed first. Research in this direction is in progress, and we hope to report the results in a future paper.

Author Contributions: Conceptualization, Methodology, and Writing—Original Draft: I.G. and H.K.T.N.; Investigation, Methodology, and Writing—review and editing: K.K. and S.W.K. All authors have read and agreed to the published version of the manuscript.

Funding: S. Kim's research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A2C1005271).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arnold, B.C.; Beaver, R.J. Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test* **2002**, *11*, 7–54. [CrossRef]
2. Arnold, B.C.; Gomez, H. Hidden truncation and additive components: Two alternative skewing paradigms. *Calcutta Stat. Assoc. Bull.* **2008**, *60*, 239–240. [CrossRef]
3. Efron, B.; Petrosian, V. Nonparametric methods for doubly truncated data. *J. Am. Stat. Assoc.* **1999**, *94*, 824–834. [CrossRef]
4. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
5. Arnold, B.C. Flexible univariate and multivariate models based on hidden truncation. *J. Stat. Plan. Inference* **2009**, *139*, 3741–3749. [CrossRef]
6. Ghosh, I.; Nadarajah, S. Inference for a hidden truncated (both-sided) bivariate Pareto (II) distribution. *Commun. Stat.—Theory Methods* **2015**, *44*, 2136–2150. [CrossRef]
7. Ghosh, I.; Ng, H.K.T. Hidden Truncation in Non-Normal Models: A Brief Survey. In *Advances in Statistics-Theory and Applications: Honoring the Contributions of Barry C. Arnold in Statistical Science*; Ghosh, I., Balakrishnan, N., Ng, H.K.T., Eds.; Springer: Cham, Switzerland, 2021; pp. 133–160.
8. Zaninetti, L. The initial mass function modeled by a left truncated beta distribution. *Astrophys. J.* **2013**, *765*, 1–7. [CrossRef]
9. Kotz, S.; Balakrishnan, N.; Johnson, N.L. *Continuous Multivariate Distributions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2000; Volume 1.
10. Ghosh, I. Bayesian inference for hidden truncation Pareto (II) models. *J. Appl. Stat. Sci.* **2014**, *20*, 257–281. [CrossRef]
11. Ghosh, I. Bayesian inference for hidden truncation Pareto (IV) models. *J. Stat. Comput. Simul.* **2020**, *90*, 2136–2155. [CrossRef]
12. Arnold, B.C.; Strauss, D. Bivariate distributions with exponential conditionals. *J. Am. Stat. Assoc.* **1988**, *83*, 522–527. [CrossRef]
13. Arnold, B.C.; Castillo, E.; Sarabia, J.M. *Conditional Specification of Statistical Models*; Springer: New York, NY, USA, 1999.
14. Gelman, A.; Hwang, J.; Vehtari, A. Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **2014**, *24*, 997–1016. [CrossRef]
15. Devroye, L. *Non-Uniform Random Variate Generation*; Springer: New York, NY, USA, 1986.
16. Murray, I.; Ghahramani, Z.; MacKay, D. MCMC for doubly-intractable distributions. In Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI-06), Cambridge, MA, USA, 13–16 July 2006; AUAI Press: Arlington, VA, USA, 2013; pp. 359–366.
17. Kass, R.E.; Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795. [CrossRef]
18. Newton, M.A.; Raftery, A.E. Approximate Bayesian inference with the weighted likelihood bootstrap. *J. R. Stat. Soc. Ser. B (Methodol.)* **1994**, *56*, 3–48. [CrossRef]
19. Moreira, C.; de Uña-Álvarez, J.; Crujeiras, R. *DTDA: Doubly Truncated Data Analysis*; R package version 3.0.1.; The Comprehensive R Archive Network: Vienna, Austria, 2022.
20. Weinberg, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*; John Wiley & Sons: New York, NY, USA, 1972.
21. Ying, Z.; Yu, W.; Zhao, Z.; Zheng, M. Regression analysis of doubly truncated data. *Astrophys. J.* **2020**, *115*, 810–821. [CrossRef] [PubMed]

22. Kundu, D.; Gupta, R.D. Bivariate generalized exponential distribution. *J. Multivar. Anal.* **2009**, *100*, 581–593. [CrossRef]
23. Marshall, A.W.; Olkin, I. A Multivariate Exponential Distribution. *J. Am. Stat. Assoc.* **1967**, *62*, 30–44. [CrossRef]