

## A Bayesian study for the comparison of generalized gamma model with its components

Bhaswati Mukherjee · Ashutosh Gupta ·  
S. K. Upadhyay

Received: 8 September 2008 / Revised: 18 January 2010 / Accepted: 9 February 2010 /  
Published online: 5 May 2011  
© Indian Statistical Institute 2011

**Abstract** Generalized gamma distribution offers a flexible family and many of the important lifetime models are obtained as component models by setting its shape parameters to unity. The flexibility of the generalized gamma model, however, occurs at the cost of its increased complexity. The present paper makes a simulation based Bayesian study to have a thorough comparison of the generalized gamma with its components in situations where the given data appear compatible with this family. Of course, if a component model is recommended the latter inferences are quite easy to deal with. The study has been conducted in two stages. First, the generalized gamma family with a scale and two shape parameters is examined to see if one or both of its shape parameters can be set to unity in order that the component models can be looked upon as possible candidates. Second, a threshold parameter added in to the family selected at the first stage is tested against zero to see if there is any desirability of threshold in the model(s). A real data set is considered for the purpose of illustration. The paper proceeds by checking compatibility of the various component models with the given data set and finally compares the models to select the one that is most pertinent with the data.

**Keywords** Generalized gamma model · Weibull model · Gamma model · Exponential model · Gibbs sampler · Predictive simulation · Partial posterior predictive  $p$ -value · BIC · DIC · Fractional Bayes factor · Posterior predictive loss

---

B. Mukherjee (✉)  
GlaxoSmithKline Pharmaceuticals Ltd., CDOCB, #5, Embassy Links, SRT Road (Cunnigham Road),  
Bangalore 560 052, India  
e-mail: bhaswati.stats@gmail.com

A. Gupta  
Biomedical Data Sciences, GlaxoSmithKline Pharmaceuticals Limited, Embassy Links 5, SRT Road,  
Bangalore 560 052, India

S. K. Upadhyay  
Department of Statistics & DST Centre for Interdisciplinary Mathematical Sciences,  
Banaras Hindu University, Varanasi 221 005, India

## 1 Introduction

Numerous parametric models have been proposed for the analysis of lifetime data. Generalized gamma model is one such flexible family often advocated to model such data sets although the model has applications in several other fields too. The model was proposed by Stacy (1962) and later on independently given by Cohen (1969). The probability density function (pdf) of the generalized gamma model in its standard form can be written as

$$f(x|\theta, \beta, \kappa) = \frac{\beta}{\Gamma(\kappa)} \frac{x^{\beta\kappa-1}}{\theta^{\beta\kappa}} \exp\left[-\left(\frac{x}{\theta}\right)^\beta\right], x > 0, \theta, \beta, \kappa > 0 \quad (1)$$

where the parameter  $\theta$  is the scale parameter and both  $\beta$  and  $\kappa$  determine the shape of the distribution. The distribution reduces to the two-parameter Weibull for  $\kappa=1$ , the two-parameter gamma for  $\beta=1$ , and the one-parameter exponential for both  $\beta=\kappa=1$ . Thus, the generalized gamma family incorporates all the important life-testing distributions and this is perhaps the reason that the model has enough scope in lifetime data analyses.

The exponential, Weibull, and gamma models, which may be referred as the component models, are all quite important in the context of lifetime data and the inferences for these are available in bulk using both classical and Bayesian methodologies. Inferential procedures for the generalized gamma distribution are, however, difficult perhaps because of an additional shape parameter. Mann et al. (1974), Lawless (1982), etc. are good references, which provide details of classical developments for these models. For Bayesian inferences, one may refer to Martz and Waller (1982), Upadhyay et al. (2001) and, more recently, Singpurwalla (2006), etc.

Obviously, the model written in the form (1) makes it more difficult to draw inferences compared to its component models although flexibility of the model (1) to deal with variety of situations cannot be denied. For the situations where (1) happens to be appropriate, the practitioners may often desire to examine the scope of its component models to be able to deal with the related inferences easily. The main purpose of our study, therefore, concentrates around providing a thorough comparison of the generalized gamma with its components when there is a dataset that appears compatible with this family. Since the entire study is Bayesian, the posterior simulation has been done either directly (in exponential case) or with the Gibbs sampler algorithm (in other cases).

Bayes paradigm offers a number of tools for studying model compatibility. We have mainly focused on the predictive simulation ideas that advocate compatibility of the data with the model if the predicted results from the model agree with the observed results (see, for example, Rubin (1984), Gelman et al. (1996) and Upadhyay et al. (2001)). It is to be noted that the predictive distribution in the present discussion is referred to mean the distribution of the future data given the observed data that may be obtained by averaging out the parameters involved in the process with respect to some of its appropriate distribution. When averaging is done with respect to the posterior, the resulting distribution may be considered as the posterior predictive distribution (see Bayarri and Berger (1998)). The simulation from the posterior predictive distribution is often difficult but MCMC approaches in

Bayes computation provide straightforward solution once the posterior samples are made available by such approaches (see, for example, Gelman et al. (1996), Upadhyay et al. (2001), etc.).

Bayesian versions of  $p$ -values based on properly chosen test statistics have also been used extensively to provide a quantitative measure for checking model compatibility through predictive simulation ideas. For details regarding such measures, readers are referred to Gelman et al. (1996), Bayarri and Berger (1998), Upadhyay et al. (2001), and Upadhyay and Mukherjee (2008), etc.

It is worthwhile to remark here that the model compatibility is related to examining a single model with the data when there is no alternative in hand. When a number of models are found compatible with the data, the practitioners are often willing to go for the most appropriate one based on their comparisons. There are several Bayesian tools available for model comparison. For the present study, we have considered the Bayesian Information Criterion (BIC), the Deviance Information Criterion (DIC) and an important version of Bayes factor, namely the fractional Bayes factor (FBF). Moreover, since model comparison involves at least two models, the basis for any such comparison should pay attention to both fitting and complexity of the models. In the same very spirit, we have also considered the posterior predictive loss (PPL) approach of Gelfand and Ghosh (1998) for comparing the models under consideration and accordingly drawing the necessary conclusions.

Before we end the section, let us have yet another important observation. It can be seen that (1) is free of threshold parameter although the addition of a threshold will further generalize the model. This generalization will lead to a non-regular family compounded with the constrained parameter region and, will further complicate the issue in a model that is already quite complicated. We have, therefore, considered the following two stages. First, we have studied the model (1) with its component models to get a single model that appears most appropriate for the data in hand. Second, we add a threshold parameter in the family chosen at the first stage and check if the threshold can be taken to be zero. If the result of comparison goes in favour of the extended family, we can justify the inclusion of a threshold parameter otherwise we consider a model chosen at the first stage without a threshold. It is to be noted that this entire strategy has been adopted to avoid the difficulty of entertaining four models each with a threshold for comparison at the first stage. As a word of remark: generalized gamma and a few of its components are difficult with a threshold but in no way it should be taken to mean that the solutions cannot be worked out as it has been successfully shown by Upadhyay and Smith (1994), Singpurwalla (2006), etc.

The plan of the paper is as under. The next section provides the model formulation, a brief review of the Gibbs sampler and its implementation details for the concerned posteriors. Section 3 considers the numerical illustration and provides full posterior analyses for all the four models based on a real data set. Section 4 relates to the model compatibility study and provides a brief review of various tools employed for the purpose. Section 5 continues with the numerical illustration for studying compatibility of all the four models with the data given in Section 3. Section 6 summarizes a brief discussion of different tools used for the comparison of the models. It is to be noted that the tools given in Section 6 are mostly borrowed from the cited literature although a description is provided to make the paper self-

content. Section 7 continues with the real data illustration and recommends a model based on the tools given in Section 6. A model chosen at this stage is then added with a threshold parameter that has been finally considered as a proposal for the second stage comparison in Section 8. Section 9 provides the numerical illustration for the second stage study using the same data set that has been considered in Section 3. The section also aims to test the appropriateness of adding a threshold parameter at the second stage. The paper finally ends with a brief conclusion given in Section 10.

## 2 Model formulation, Gibbs sampler, and the posterior exploration

Gibbs sampler provides a way for generating samples from the posteriors that are specified up to proportionality only. The scheme proceeds iteratively by generating variate values in a cyclic manner from the various full conditionals, specified up to proportionality, using suitably chosen initial values for starting the process. The iterations are carried out until a systematic pattern of convergence is achieved through the generating scheme. It has been observed that after a large number of iterations the generated samples converge in distribution to a random sample from the true posterior distribution (see, for example, Smith and Roberts (1993), Upadhyay et al. (2001) and Robert and Casella (2004), etc. for a detailed discussion).

To provide the implementation details of the algorithm for the generalized gamma, let us assume that  $\underline{x}$ :  $x_1, x_2, \dots, x_n$  be the observed lifetimes from (1). The corresponding likelihood function can be written as

$$L(\underline{x}|\theta, \beta, \kappa) = \left(\frac{\beta}{\Gamma(\kappa)\theta^{\beta\kappa}}\right)^n \prod_{i=1}^n (x_i)^{\beta\kappa-1} \exp\left[-\sum_{i=1}^n \left(\frac{x_i}{\theta}\right)^\beta\right]. \tag{2}$$

We consider independent prior for  $\theta$ ,  $\beta$  and  $\kappa$  (see Upadhyay et al. (2001)) given as

$$\left. \begin{aligned} g_1(\theta) &\propto \frac{1}{\theta} \\ g_2(\beta) &\propto G(a_1, b_1) \\ g_3(\kappa) &\propto G(a_2, b_2) \end{aligned} \right\} \tag{3}$$

where  $G(a, b)$  stands for the gamma distribution with shape parameter  $a$  and scale parameter  $b$ . It should be noted here that except for the prior distribution of  $\theta$ , the prior distributions for rest of the parameters are proper. Moreover, the prior for  $\theta$  is taken as per Jeffreys’ suggestion for the priors of the scale parameters. The above choices, therefore, ensure the posterior to be proper (see also Upadhyay et al. (2001), Singpurwalla (2006)). Obviously, the posterior up to proportionality can be written as

$$p(\theta, \beta, \kappa|\underline{x}) \propto \frac{\beta^{n+a_1-1} \kappa^{a_2-1}}{(\Gamma(\kappa))^n \theta^{n\beta\kappa+1}} \prod_{i=1}^n (x_i)^{\beta\kappa-1} \exp\left[-\left(\sum_{i=1}^n \left(\frac{x_i}{\theta}\right)^\beta + \frac{\beta}{b_1} + \frac{\kappa}{b_2}\right)\right]. \tag{4}$$

The full conditionals of  $\theta$ ,  $\beta$  and  $\kappa$  can be easily made available from (4). Upadhyay et al. (2001) have provided the necessary details on sample generation schemes from these various full conditionals (see also Singpurwalla (2006)). Thus, the Gibbs sampler algorithm can be easily applied to model (1) to simulate the corresponding posterior.

For implementation of the Gibbs sampler algorithm to models such as the Weibull and/or the gamma, one can similarly proceed by putting  $\kappa=1$  and/or  $\beta=1$  in the likelihood (2) and omit the corresponding prior and the variate in (4). Thus, the implementation of the algorithm to these component models will be even more simplified. The exponential model ( $\beta=1$ ,  $\kappa=1$ ) is perhaps the easiest one and if we consider the same prior for  $\theta$  as in (3), the posterior simplifies to

$$p(\theta|\underline{x}) \propto \frac{1}{\theta^{n+1}} \exp\left(-\sum_{i=1}^n \frac{x_i}{\theta}\right), \quad (5)$$

which is a gamma distribution and it can be directly simulated using any gamma generating routine (see, for example, Devroye (1986)).

### 3 Numerical illustration: posterior analysis based on a real data

The data set considered for the purpose of numerical illustration refers to number of million revolutions before failure for each of 23 bearings in a life test experiment. The complete set of observations is shown in Table 1. Lieblein and Zelen (1956) analyzed the data for Weibull model whereas Lawless analyzed it for both Weibull and generalized gamma models. Upadhyay et al. (2001) considered this data in the context of generalized gamma model while Bain and Engelhardt (1980) used it for comparing Weibull and gamma models. It, therefore, appears that not only the generalized gamma but also a few of its important components appear to be strong contenders for this data set. It is to be noted that if one goes for recommending the component models, one is actually saved from the inherent complexities of the generalized gamma.

Assuming the data set from each of the four models, in turn, we analyzed the corresponding posteriors as per details provided in the previous section taking  $a_i=2.00$  and  $b_i=1.00$  ( $i=1, 2$ ). The values 2.00 and 1.00 were simply taken following Upadhyay et al. (2001) although variations in priors because of the changes in hyperparameters (or even the modeling assumptions) do not make any detrimental consequence as long as the tools such as those of sampling-importance-resampling are available (see, for example, Upadhyay and Smith (1993)). Moreover, since the objective of the paper is model comparison in a Bayesian framework, the paradigm checks the entire modeling assumption that includes the prior as well.

**Table 1** Million of revolutions before failure for each of 23 ball bearings

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

A few of the important results of our posterior analyses in the form of posterior modes and means are shown in Table 2. These results are based on a sample of size 1,000 from each of the four posteriors obtained either directly in exponential case or through the proper implementation of the Gibbs sampler algorithm in other cases. We skip any discussion on these results simply because they represent the usual point estimates; posterior mode being the most probable value whereas posterior mean being the optimal estimate under squared error loss function.

#### 4 Model compatibility

There are several tools for studying model compatibility. An important and logically convincing idea to check model compatibility is based on predictive simulation, which suggests that a model is compatible if it provides the prediction in accordance with the patterns given in the observed data  $\underline{x}$ . In case of no or poor resemblance of the predicted data with the observed data, the model is suspected. This resemblance can often be measured on the basis of a suitable test statistic and the result can be presented either graphically or by means of a quantitative measure such as the  $p$ -value (see, for example, Gelman et al. (1996), Bayarri and Berger (1998), Upadhyay and Peshwani (2003), etc.).

For the present study, we consider the empirical distribution function (Edf) plots for the observed and the predicted data from the model(s) (see Upadhyay et al. (2001)) as well as some other plots based on some future order statistics that we feel important for model discrimination in life testing studies. These plots often provide an informal but a striking message on model compatibility. To provide quantitative evidence on the study of model compatibility, we advocate for the use of an important Bayesian version of  $p$ -value based on some statistics with large values of the latter casting doubt on the model authenticity.

To discuss briefly the notion of  $p$ -value, let us consider the data  $\underline{x}$  from an assumed model  $f(\cdot)$  given under the hypothesis  $H_0: X \sim f(x|\Theta)$  where  $\Theta$  stands for the unknown model parameter(s). Suppose further that we have a test statistic  $T(X)$  to investigate compatibility of the model with the observed data. The  $p$ -value is then defined as

$$p = P[T(X) \geq t(\underline{x})] \quad (6)$$

where  $t(\underline{x})$  denotes the value of  $T(X)$  at the observed data  $\underline{x}$ . Obviously, the large value of  $t(\underline{x})$  cast doubt on compatibility of the model with the given data as it increases the chance of  $p$  getting smaller.

**Table 2** Posterior modes and means for different parameters under the assumption of each of the four models

Models	Posterior modes			Posterior means		
	$\theta$	$\beta$	$\kappa$	$\theta$	$\beta$	$\kappa$
Exponential	67.634	–	–	75.235	–	–
Weibull	83.410	2.036	–	83.054	2.057	–
Gamma	20.691	–	3.393	23.081	–	3.408
Generalized gamma	42.151	1.291	1.697	50.019	1.531	2.068

The main problem with the  $p$ -value defined in (6) is its dependence on the unknown parameter  $\Theta$ . A number of suggestions have been made both in the classical as well as in the Bayesian literature to get rid of this unknown parameter but we are not going into the details of these various suggestions. The interested readers may refer to Bayarri and Berger (1998) and, more recently, Upadhyay and Mukherjee (2008) for details. It is to be noted that the main logic behind Bayesian version of  $p$ -value is to average out the unknown  $\Theta$  with respect to any of its suitable distribution.

The present paper focuses on partial posterior predictive  $p$ -value (PV) suggested by Bayarri and Berger (1998) mainly because of its advantages over other versions and perhaps because of its comparative ease of computation. PV can be defined as

$$P = \int \text{pr}[T(X) \geq t(\underline{x})] \cdot p^*(\Theta) \cdot d\Theta \tag{7}$$

where  $p^*(\Theta)$  is the partial posterior given as

$$p^*(\Theta) \propto L(\underline{x}|t(\underline{x}), \Theta) \cdot g(\Theta) \propto \frac{L(\underline{x}|\Theta) \cdot g(\Theta)}{L(t(\underline{x})|\Theta)}.$$

For evaluating (7) one needs repeated generation from  $p^*(\Theta)$  and then correspondingly  $T$  from  $f(t|\Theta)$ , the latter can, of course, be done by generating  $X$  from  $f(x|\Theta)$  and computing  $T$ . The fraction of generated  $T$  that exceeds  $t(\cdot)$  will then provide the estimate of the  $p$ -value. It is to be noted that generation of  $X$  from  $f(x|\Theta)$  when  $\Theta$  has already been obtained from  $p^*(\Theta)$  can be regarded in a predictive sense and, therefore, a better notation can be  $Y$  instead of  $X$ . We are, however, writing  $X$  to maintain notational uniformity that has been given in (6).

A possible way for generating  $\Theta$  from  $p^*(\Theta)$  may include the use of Metropolis chain where the full posterior density  $p(\Theta|x)$  may itself be taken as the probing distribution. Alternatively, PV can also be computed approximately by using an importance sampling based estimate given as

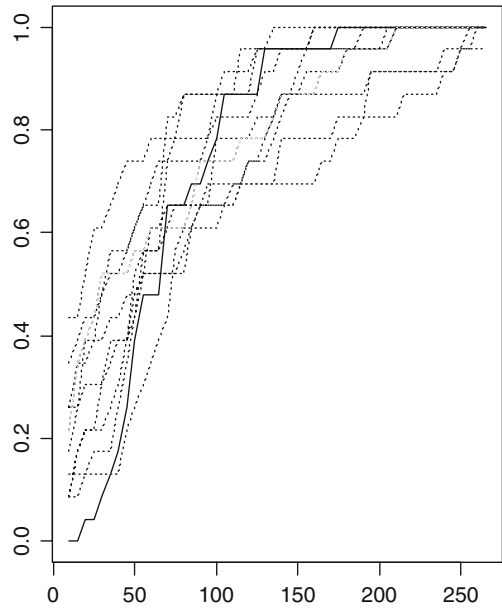
$$\hat{PV} = \frac{\sum_{i=1}^m \text{Pr}(T \geq t(\cdot)|\Theta_i) / f(t(\cdot)|\Theta_i)}{\sum_{i=1}^m 1 / f(t(\cdot)|\Theta_i)}, \tag{8}$$

where  $\Theta_1, \Theta_2, \dots, \Theta_m$  constitute a sample of size  $m$  obtained from the posterior  $p(\Theta|\underline{x})$ . Readers may also refer to Bayarri and Berger (1998) for the details of various computational strategies.

**5 Numerical illustration continued: examining model compatibility based on real data**

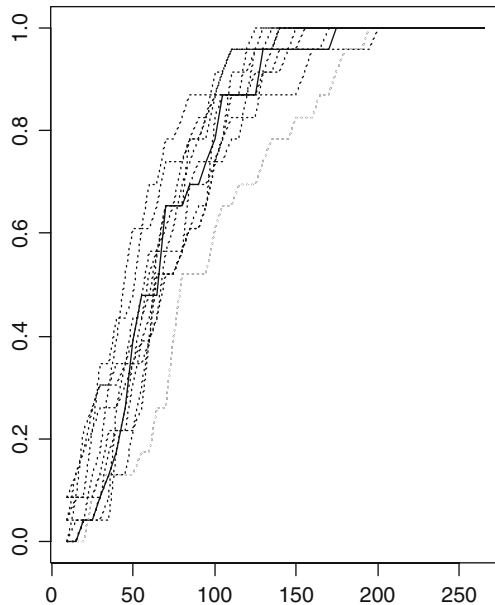
For examining model compatibility, we considered the same data set reported earlier in Table 1. First of all, we plotted the Edfs for the observed and the predicted data on the same scale corresponding to the generalized gamma and its components. For this purpose, we generated ten predictive datasets each of size equal to that of the observed dataset from each of the considered models separately using the

**Fig. 1** Edf plots for the observed and the predictive data corresponding to the exponential model (*continuous line* corresponds to observed data based Edf)



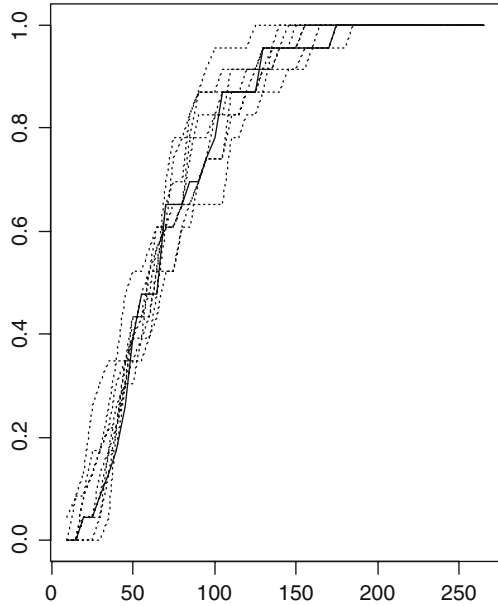
corresponding posterior samples. The plots are shown in the Figs. 1, 2, 3, and 4 with broken lines corresponding to the predictive data. We observe that the figures are quite supportive to at least Weibull, gamma and generalized gamma models. The exponential model may be slightly indifferent in the tails but it also cannot be rejected at a first glance.

**Fig. 2** Edf plots for the observed and the predictive data corresponding to the Weibull model (*continuous line* corresponds to observed data based Edf)



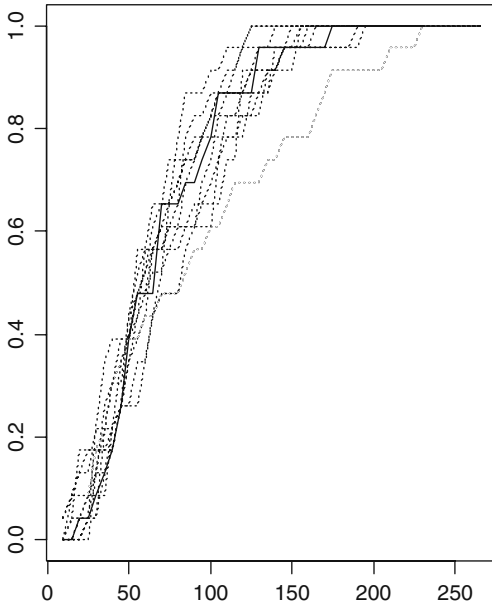


**Fig. 3** Edf plots for the observed and the predictive data corresponding to the gamma model (*continuous line* corresponds to observed data based Edf)

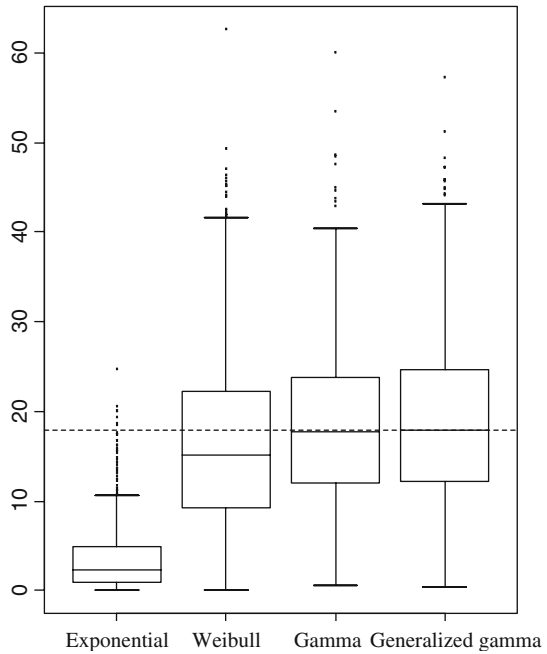


To examine our conclusion especially in the tail areas, we next consider some additional plots based on the extreme observations. Figures 5 and 6 show the density estimates of lowest and highest order future statistics ( $Y_1$  and  $Y_n$ ) based on predictive samples of size 1,000 from the corresponding models. The dashed lines in the figures represent the observed values of the same order. It can be seen here too

**Fig. 4** Edf plots for the observed and the predictive data corresponding to the generalized gamma model (*continuous line* corresponds to observed data based Edf)

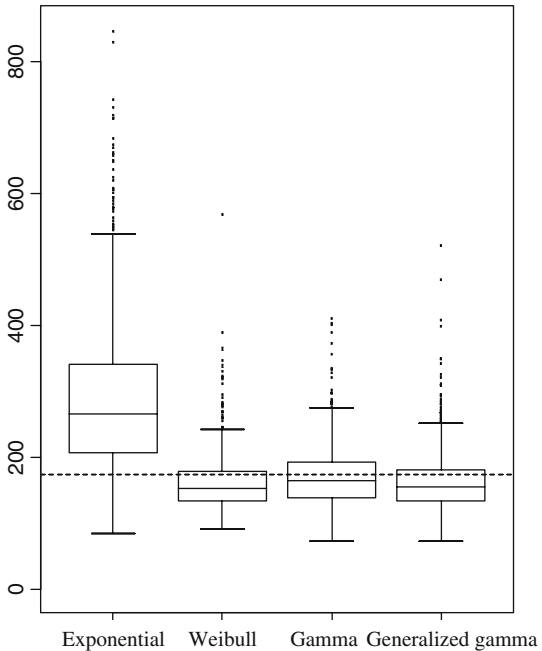


**Fig. 5** Box plots showing the density estimates of  $Y_1$  under the four models (*dashed line* corresponds to the smallest observed data)



that both the extreme observations are not supportive to the exponential model the way they are supporting the other models. The smallest observed value rather conveys to go against the exponential assumption.

**Fig. 6** Box plots showing the density estimates of  $Y_n$  under the four models (*dashed line* corresponds to the largest observed data)



We finally evaluated  $\hat{PV}$  based on a few smallest (for example,  $Y_1, Y_2,$  and  $Y_3$ ), a few largest (for example,  $Y_{n-1}, Y_n$ ) and a few middle (say,  $Y_{Md-1}, Y_{Md}, Y_{Md+1}$ ) ordered predictive test statistics. These statistics will provide an idea about the overall compatibility of the models with the observed data. Whereas the smallest and the largest ordered observations will check the compatibility at the lower and upper tails, the middle ordered observations would reflect the information about the central region of the hypothesized model. It is to be noted that the predictive sample size is same as the observed sample size and, therefore, the largest order test statistic corresponds to  $n=23$  whereas the middle order test statistic corresponds to  $Md=12$ . Also,  $Md=12$  is an arbitrary selection which simply means a value in the central part of the predictive data.

Table 3 summarizes the estimated PV for various test statistics assuming each of the four models separately. The values are importance-sampling based estimates of PV as given in (8). It can be further noted that while calculating PV, the prior for the scale parameter for all the four models was taken to be proportional to unity. This has been done to ensure that the corresponding partial posteriors remain proper (see, for example, Upadhyay and Mukherjee (2008)).

Obviously, the  $p$ -values corresponding to lower ordered test statistics are not supportive to the exponential distribution, a finding that changes with regard to higher ordered test statistics where all the four models appear compatible and none can be rejected based on these measures. On the other hand, if we consider the other three models they are found to provide good fit with the data whatever order of predictive statistics we choose. Thus, our model compatibility study conveys that at least three of the models, namely Weibull, gamma, and generalized gamma, are good candidates for the data in hand. The situation can be somewhat regarded under the context of Bayesian robustness analysis where one can claim that the generalized gamma distribution and its two components are insensitive to have a narrow discriminatory capability. The normal practice may be try using parsimony principle to select the simplest model but we shall stick to a thorough comparison of these models before we draw any final conclusion in favour of one among these. As a word of final comment, it should be made clear that the model compatibility study given above should not be taken as a deciding factor to go against a model and hence we propose to include exponential model as well in the said comparison.

**Table 3**  $\hat{PV}$  for the four models based on a few selected ordered future observations ( $Md=12, n=23$ )

Model	$\hat{PV}$ based on the test statistic							
	$Y_1$	$Y_2$	$Y_3$	$Y_{Md-1}$	$Y_{Md}$	$Y_{Md+1}$	$Y_{n-1}$	$Y_n$
Exponential	0.004	0.003	0.004	0.252	0.211	0.252	0.185	0.187
Weibull	0.300	0.254	0.312	0.249	0.285	0.255	0.282	0.280
Gamma	0.469	0.230	0.309	0.248	0.301	0.255	0.401	0.439
Generalized gamma	0.488	0.252	0.314	0.250	0.311	0.258	0.391	0.367

## 6 Model comparison

The following subsections sketch briefly the different model comparison criteria that have been taken into account. Readers may refer to cited references for further details.

### 6.1 Bayes and deviance information criteria

A good model comparison criterion suggests a model that incorporates the main features of the data and not the noise involved therein and, therefore, we advocate for the use of a criterion which includes terms for both model fit and model complexity. Two most common selection criteria, which include both complexity and goodness of fit, are BIC (see Schwarz (1978)) and DIC (see Spiegelhalter et al. (2002)).

The BIC chooses a model that minimizes

$$\text{BIC} = -2 \log L(\underline{x} | \hat{\theta}) + k \ln(n), \quad (9)$$

where  $L(\underline{x} | \hat{\theta})$  is the maximum likelihood (ML) corresponding to the model under consideration and  $k$  is the number of components in  $\theta$ . The first term in (9) corresponds to model fit and the second is a penalty term which promotes model parsimony by penalizing the model that is more complex. The BIC is a consistent measure in the sense that it provides the probability of selecting the correct model tending towards unity as the number of observations approaches to infinity. It also has advantages of simplicity and (an apparent) freedom from prior assumptions.

For computation of BIC, we consider a slight approximation and use posterior mode in place of ML estimator in (9). This has been done partially because of ease and partially because of maintaining uniformity with all the models as ML estimator may often fail to exist with the generalized gamma (see, for example, Lawless (1982)). Moreover, since the priors are not too strong, we feel as if we do not deviate much from the BIC based on ML estimators. It is to be noted that similar approximation in the evaluation of BIC has also been suggested by Sahu and Dey (2000) among others.

DIC was introduced by Spiegelhalter et al. (2002). The formulation of DIC is based on examining the posterior distribution of the classical deviance (see also Dempster (1974)), which is defined as

$$D(\theta) = -2 \log L(\underline{x} | \theta) + 2 \log h(\underline{x}) \quad (10)$$

where  $\log h(\underline{x})$  denotes a fully specified standardizing term and is a function of data alone. The DIC, a measure based on the posterior distribution of  $D(\theta)$ , can then be defined as

$$\text{DIC} = \overline{D(\theta)} + p_D \quad (11)$$

where  $p_D = \overline{D(\theta)} - D(\tilde{\theta})$  and is known as effective number of parameters. It can also be thought of as a measure of model complexity.  $\tilde{\theta}$  is an estimate of  $\theta$  which depends upon  $x$  only. Spiegelhalter et al. (2002) recommended choosing posterior mean as an estimate of  $\theta$ . The first term of Eq. 11 is the posterior expectation of the

deviance and it can be considered as a Bayesian measure of model fit. Thus, DIC provides a model comparison criterion, which depends on both the goodness of fit and complexity of the models. A model, giving the smallest value of DIC among all the competing models, is finally recommended.

## 6.2 Fractional Bayes factor

To provide a formal definition of Bayes factor, consider the two models or the hypotheses being specified as  $H_j$ : data  $\underline{x}$  support model  $M_j = f_j(\underline{x}|\theta_j)$ ,  $j=1, 2$ ; with unknown parameter vector  $\theta_j$  associated with  $M_j$ . The Bayes factor for  $M_2$  against  $M_1$  can be defined as

$$B_{21} = \frac{\int L_2(\underline{x}|\theta_2)g_2(\theta_2)d\theta_2}{\int L_1(\underline{x}|\theta_1)g_1(\theta_1)d\theta_1} \quad (12)$$

where  $L_j(\cdot)$  is the likelihood function corresponding to  $M_j$  and  $g_j(\theta_j)$  is the prior distribution of  $\theta_j$ . Thus the unknown parameter vector  $\theta_j$  can be successfully eliminated from (12) by averaging the likelihood with respect to the prior distribution of  $\theta_j$ .

The Bayes factor defined in (12) appears simple but suffers from a major caveat when the priors are non-informative and defined up to arbitrary constants (see, for example, Berger and Pericchi (1996)). There have been several suggestions in the literature to deal with such situations. We, however, confine our discussion to FBF introduced by O'Hagan (1995) mainly because of its inherent advantages. The FBF considers the likelihood in two parts, one corresponding to (usually) a small fraction and the other for the remaining fraction. The small fraction of the likelihood is used to convert the non-informative (usually improper) prior to a proper posterior. This proper posterior is then used to obtain the posterior mean of the major fraction of the likelihood (see O'Hagan (1995) for other details). Thus FBF can be defined as,

$$B_{21}^F = \frac{\int L_2^{1-b}(\underline{x}|\theta_2) p_2^b(\theta_2|\underline{x})d\theta_2}{\int L_1^{1-b}(\underline{x}|\theta_1) p_1^b(\theta_1|\underline{x})d\theta_1} \quad (13)$$

where  $p_j^b(\theta_j|\underline{x})$  denotes the posterior corresponding to the small fraction  $b$  of the likelihood. O'Hagan (1995) has provided some suggestions on the choices of  $b$  that vary according to the degree of robustness desired by the experimenter.

## 6.3 Posterior predictive loss approach

We consider an additional tool for comparing the models that facilitates in decision-making too. The tool for model comparison based on predictive distributions was proposed by Gelfand and Ghosh (1998) where the authors advocated to minimize loss over the concerned models following the standard utility ideas discussed in Raiffa and Schlaifer (1961). To clarify, let us suppose that  $\underline{x}$  is the vector of observed values with components  $x_i$  ( $i=1,2,\dots,n$ ). Also suppose that  $\underline{y}$  with components  $y_i$  denote the future set of observations from the same model under consideration. We further assume the action vector  $a$ , which is an estimate trying to accommodate both

$\underline{x}$  and  $\underline{y}$ , the latter being the quantity that we predict for  $\underline{x}$ . Finally, suppose that  $L(\underline{y}, a; \underline{x})$  denotes the loss incurred in taking action  $a$  when  $\underline{y}$  obtains and  $\underline{x}$  was observed. The authors finally suggested choosing a model that minimizes the expected loss over  $a$  where expectation is taken with respect to the posterior predictive distribution of  $\underline{y}$ . For the  $i$ th component of  $\underline{y}$  and the action vector  $a$ , the authors partitioned the loss function as

$$L(y_i, a_i; \underline{x}) = L(y_i, a_i) + kL(x_i, a_i), \quad (14)$$

where the action  $a_i$  tries to accommodate both  $x_i$  and  $y_i$ . This way  $a_i$  is a univariate compromising action for the partially unknown bivariate state of nature  $(x_i; y_i)$ . The predetermined constant  $k$  ( $\geq 0$ ) is the weight that indicates the relative regret for the departure from  $x_i$  compared with departure from  $y_i$ . The first term on the right hand side of Eq. 14 gives the precision of estimation and the second term is a measure for goodness of fit. Thus the criterion proposed by Gelfand and Ghosh (1998) combines both the predictive variability of the model and its performance at the observed data points.

If (14) is aggregated over the components of  $\underline{y}$ , the expression for minimum expected loss function corresponding to a model, say  $m$ , could be given as

$$D_k(m) = \sum_{i=1}^m \min_{a_i} E_{y_i|\underline{x}, m} L(y_i, a_i; \underline{x}). \quad (15)$$

For the squared error loss function, often used for its inherent simplicity, the minimization can be carried out explicitly (see, for example, Sahu and Dey (2000)). The expression for the squared error loss function can be written as

$$D_k(m) = \frac{k}{k+1} G(m) + P(m) \quad (16)$$

where  $G(m) = \sum_{i=1}^n (\mu_i^m - x_i)^2$ ,  $P(m) = \sum_{i=1}^n \sigma_i^{2(m)}$ ,  $\mu_i^{(m)} = E(y_i|\underline{x}, m)$  and  $\sigma_i^{2(m)} = \text{var}(y_i|\underline{x}, m)$ .

A simplified version of (16), that becomes independent of  $k$ , was given by Laud and Ibrahim (1995) and later on used by Sahu and Dey (2000) in the context of analyzing bivariate survival data. In fact, the version proposed by Gelfand and Ghosh (1998) can be considered as the generalized version of the criterion proposed by these authors.

## 7 Numerical illustration continued: model comparison and choice based on real data

We compared the models, for the data given in Table 1, based on the tools discussed in Section 6. The BIC and DIC values evaluated for each model are shown in Table 4. Throughout we have considered posterior samples of size 1,000 corresponding to the generalized gamma and its component models as per the discussion given in Section 2.

It is obvious from the table that BIC and DIC values for the gamma model are the smallest though the values corresponding to the generalized gamma and Weibull

**Table 4** BIC and DIC values for generalized gamma and its components based on the observed data

Model	BIC	DIC
Exponential	246.104	244.985
Weibull	233.756	228.984
Gamma	232.738	226.653
Generalized gamma	239.350	230.419

models are not much apart. Therefore, it appears that all the three models, namely Weibull, gamma and generalized gamma, appear to be good contenders though gamma seems to be the most appropriate choice. Exponential model, on the other hand, is certainly the weakest candidate for the data and one may not be advised to entertain it. Thus, if one has to provide a preference pattern for the models based on these two measures, the order will be gamma, Weibull, generalized gamma, and lastly the exponential. Before sticking to this as a final conclusion, let us examine other tools as well.

The values of FBF are shown in the Table 5 for all the possible pairs of models. In the table, scripts 1, 2, 3, and 4 are used to denote generalized gamma, gamma, Weibull, and exponential models, respectively. In addition,  $B_{ji}^F$  stands for FBF, for model  $j$  against model  $i$ . It is to be noted here that the expressions of FBF for different pairs of models are not available in closed form and, therefore, we relied on simulation-based techniques, especially the one using Gibbs sampler algorithm. For the purpose we employed the strategies detailed in Section 2 but confined to the fraction  $b$  of various posteriors. We finally used Eq. 13 to find out the means of remaining fraction of likelihoods using the first stage posteriors. The value of  $b$  was taken to be  $0.20 (\approx \sqrt{n/n})$  in each case following a suggestion given by O’Hagan (1995).

Table 5 shows that the values of FBF are equally favouring the generalized gamma and the gamma models for the data in hand. The Weibull model also does not appear to be a bad choice although the preferential ordering based on evaluated FBF certainly recommends the generalized gamma or the gamma model. In any case, the exponential model is not to be considered at all. Thus, we came across the same preferential pattern for selecting the models that was given earlier on the basis of BIC and DIC. One nice property with FBF is that this measure is coherent. This means that for the three models labeled as  $M_1, M_2$  and  $M_3$ , if FBF is computed pair wise then the product of the FBF of  $M_1$  vs.  $M_2$  and  $M_2$  vs.  $M_3$  is equal to the FBF of  $M_1$  vs.  $M_3$ . It is to be noted that our reported results do not satisfy this coherency

**Table 5** FBF for model  $j$  ( $=1, 2, 3$ ) against  $i$  ( $=2, 3, 4$ ) based on the observed data

Values of (j, i)	$B_{ji}^F$
(1, 2)	1.015
(1, 3)	2.935
(1, 4)	361.872
(2, 3)	2.916
(2, 4)	348.370
(3, 4)	176.531

property exactly. This may be because of the fact that a few of the considered models are complicated and the computational strategies based on simulation can never be exact. Thus the values depicted in Table 5 may be regarded as estimates of FBF in a true sense.

We finally calculated the PPL for the models under consideration using the squared-error loss function. The values are shown in Table 6. The table clearly shows that the overall loss corresponding to the gamma model is least. If, however, we look on other values, we find that the losses corresponding to the generalized gamma and the Weibull models are not much larger than that of the gamma model. For the generalized gamma, the obvious reason is that while the model is not providing appreciable gain in fitting over gamma, it is being penalized enough due to its inherent complexity and thereby increasing the overall loss. For the Weibull model, it was seen that its fit was not as good as that of the other two models and this fact is reflected here too in terms of the losses due to fitting. Its complicity factor is, however, small in comparison to the generalized gamma though not appreciably penalized with respect to gamma, a fact that can be clearly seen in the table. It must be noted that the loss due to penalty is nothing but the predictive variance and hence it inflates both for simple and overly complicated models. The exponential model is simplest and, therefore, giving the largest penalty loss and loss due to fitting which result in overall loss much larger. Thus, our decision is to choose gamma or at most the Weibull model for the given data

### 8 Three-parameter gamma model—a proposal for second stage comparison

In the preceding sections, we compared the generalized gamma model with its components and observed that gamma model happens to be an appropriate choice for the entertained data given in Table 1. The generalized gamma and all its components were considered without a threshold parameter to make the study simpler. As mentioned in Section 1, we now incorporate a threshold parameter in the two-parameter gamma model, a model selected through an extensive study given in the previous section. We then test this threshold parameter against zero to see if the two-parameter gamma model is really justified for the entertained data or there is some scope of a threshold hidden in the model. This is worth mentioning that we are doing this only for the sake of simplicity and it does not always ensure that the best of the possible combinations, with or without a threshold, is finally selected. Thus, to begin our second stage model comparison study, we notice that the three-

**Table 6** PPL for the four models based on observed data

Model	Loss due to penalty	Loss due to fitting	Overall loss
Exponential	5.859	55.868	56.649
Weibull	2.018	41.965	40.168
Gamma	1.676	40.550	38.540
Generalized gamma	7.587	41.253	45.089



parameter gamma model can be easily accessed by putting  $\beta=1$  and replacing  $x$  with  $(x-\mu)$ ,  $x > \mu$ , in (1). The corresponding likelihood function can be written from (2) as

$$L(\underline{x}|\theta, \kappa, \mu) = \left(\frac{1}{\Gamma(\kappa)\theta^\kappa}\right)^n \prod_{i=1}^n (x_i - \mu)^{\kappa-1} \exp\left[-\sum_{i=1}^n \left(\frac{x_i - \mu}{\theta}\right)\right]. \quad (17)$$

If we consider the same prior for  $\kappa$  and  $\theta$  as given in (3) and allow the threshold to have a prior proportional to constant, the posterior up to proportionality can be written as

$$p(\theta, \kappa, \mu|\underline{x}) \propto \frac{\kappa^{a_2-1}}{(\Gamma(\kappa))^n \theta^{n\kappa+1}} \prod_{i=1}^n (x_i - \mu)^{\kappa-1} \exp\left[-\left(\sum_{i=1}^n \left(\frac{x_i - \mu}{\theta}\right) + \frac{\kappa}{b_2}\right)\right]. \quad (18)$$

The posterior (18) can be an easy candidate for the Gibbs sampler implementation as all its full conditionals are available from the viewpoint of sample generation. The details of the various full conditionals and related sample generation schemes can be had from Upadhyay and Smith (1994) as a special case of their entertained model formulation.

## 9 Numerical illustration for second stage study

### 9.1 Posterior analysis for the second stage study

For numerical illustration, we proceeded with a single Gibbs chain of long run and, after the convergence monitoring was assessed, took equally spaced (every 10th) outcome to form the samples from the posterior (18). The prior hyperparameter and other implementation strategies were same as those given in Section 2.

Table 7 provides the posterior modes and means though other characteristics can be likewise drawn once the posterior samples are obtained. These results are based on a sample of size 1,000. The values are self-explanatory and, therefore, do not require any discussion in the light of our primary objective of model comparison.

### 9.2 Compatibility study of the three-parameter gamma model

As mentioned, the comparison of two models for a data is justified if both happen to provide good compatibility with the data. We have already seen the compatibility of

**Table 7** Posterior modes and means for different parameters under the assumption of the three-parameter gamma model

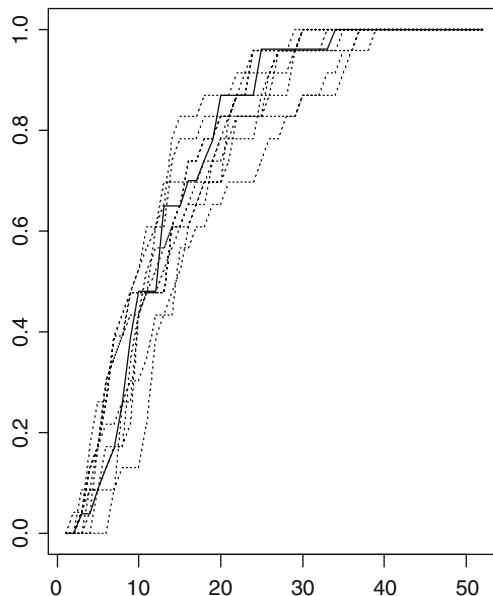
Model	Posterior modes			Posterior means		
	$\theta$	$\kappa$	$\mu$	$\theta$	$\kappa$	$\mu$
Three-parameter gamma	21.236	2.473	10.301	25.651	2.776	8.854

two-parameter gamma model for the data given in Table 1 (see Section 4) so we are left only with the compatibility study of the three-parameter gamma model for the same data. If the latter compatibility is established, we shall proceed for model comparison otherwise our conclusion of model choice will remain in favour of two-parameter gamma model.

The study begins with Edf plots for both the predicted and the observed data under the assumption of the three-parameter gamma model. Figure 7 shows the corresponding Edf plots based on the observed data and the ten predicted data, each of size similar to that of the observed data. One can clearly see that the model provides a good compatibility with the data in hand (see Fig. 7). Moreover, since the study based on these plots is an informal one, we also considered the evaluation of PV as per the details given in Section 4. The estimated PVs based on the same future order statistics (see also Section 5) are shown in Table 8. It is obvious that the estimates are quite supportive to the three-parameter gamma model as far as the compatibility with the observed data is concerned.

The compatibility of the three-parameter gamma model was further examined, especially in the tail area regions by studying the density estimates of the lowest and the highest order future observations based on samples of size 1,000 from the model. The density estimates are shown in the Fig. 8. In the figure  $GM_3$ -Min(Max) denotes the plot corresponding to smallest(largest) ordered future observation from the three-parameter gamma model. The dashed lines in the figure represent the corresponding observed values. It is obvious from the figure that the observed highest and lowest data points appear quite probable in the distributions of the corresponding predictive statistics giving us a clear-cut impression about the model compatibility in the extreme tails, the region where the lifetime distributions often mismatch and provide disagreement with the data.

**Fig. 7** Edf plots for the observed and the predictive data for the three-parameter gamma model (*continuous line* corresponds to observed data based Edf)



**Table 8**  $\hat{P}V$  for the three-parameter gamma model based on a few selected ordered future observations ( $Md=12, n=23$ )

Model	$\hat{P}V$ based on the test statistic							
	$Y_1$	$Y_2$	$Y_3$	$Y_{Md-1}$	$Y_{Md}$	$Y_{Md+1}$	$Y_{n-1}$	$Y_n$
Three-parameter gamma	0.348	0.503	0.513	0.473	0.486	0.504	0.497	0.562

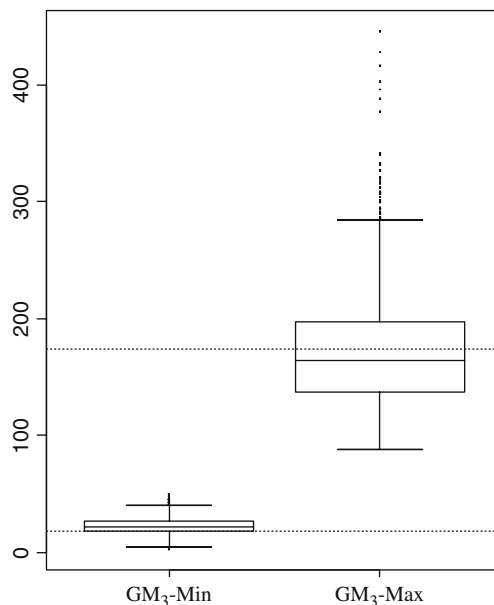
### 9.3 Model comparison: gamma with threshold vs. gamma without threshold

The model compatibility study given in Sections 5 and 9.2 are supportive to both two-parameter and three-parameter gamma models and hence the comparison of these models can be taken up to see if the threshold parameter is really desired for the data in hand. We may, therefore, proceed for the said comparison using the tools discussed in Section 6.

The evaluated values of BIC and DIC are shown in Table 9. The values corresponding to the two-parameter form are borrowed from Table 4 and these have been rewritten for the purpose of comparison only. One can clearly see that the differences between the BIC and DIC values corresponding to the two models are not significant and, therefore, one can advocate the use of simpler two-parameter gamma model. This is otherwise conveyed too by the parsimony principle.

As it has been done earlier, we do not summarize our conclusion just based on the above two information criteria rather see the message obtained through the evaluation of FBF and PPL. It is to be noted here that the three-parameter gamma model is an irregular constrained parameter family and, as such, FBF always tends to

**Fig. 8** Box plots showing the density estimates of  $Y_1$  and  $Y_n$  for the three-parameter gamma model (*dashed lines* correspond to the smallest and largest observed data)



**Table 9** BIC and DIC values for the two-parameter and the three-parameter gamma models based on the observed data

Model	BIC	DIC
Two-parameter gamma	232.738	226.653
Three-parameter gamma	233.084	225.988

favour a more complex model. Thus following Berger and Pericchi (2001), we shall refrain from using FBF and draw our conclusion based on only PPL, the details of which are provided in subsection 6.3 (see also Laud and Ibrahim (1995)).

The values of PPL for the two gamma models are shown in Table 10. The values corresponding to the two-parameter form of the model are borrowed from Table 6 and are reproduced here for an easy impact of comparison. It is apparent from the table that the overall loss corresponding to the three-parameter gamma model is smaller than that of the two-parameter form. This clearly gives an indication of threshold in the model so that no important information supplied by the data goes astray. We also notice that the loss due to penalty associated with two-parameter model is higher than the corresponding value for the three-parameter form. This loss is nothing but the predictive variance which may sometimes increase with the oversimplified model (see Gelfand and Ghosh (1998)) thereby giving higher loss due to penalty. This finding is a bit surprising as it conveys that the two-parameter gamma model is an over simplified model for this data set.

## 10 Conclusion

Model comparison or model choice has always been a challenging task but thanks to the availability of various tools, a few of them successfully described and explored in the present paper. After a thorough comparison of the generalized gamma with its components, we observe that one should not simply follow the flexibility or parsimony principle but also look on various aspects before recommending a model. As a result, we have neither gone in favour of exponential nor in favour of generalized gamma but suggest the use of gamma model for the considered data set. Our entire study has been undertaken in two stages. The first stage entertains each model without threshold whereas the second, after getting a model from the first stage, incorporates a threshold to test for its authenticity. This way we have avoided, at first stage, the intricacy of entertaining the generalized gamma and all its components with threshold but finally recommended a model with threshold parameter.

**Table 10** PPL for the two gamma models based on the observed data

Model	Loss due to penalty	Loss due to fitting	Overall loss
Two-parameter gamma	1.676	40.550	38.540
Three-parameter gamma	1.182	40.477	37.981

**Acknowledgement** The authors express their thankfulness to the referees and the Editor for providing some concrete suggestions that helped in improving the earlier draft.

## References

- Bain, L.J., and M. Engelhardt. 1980. Probability of correct selection of Weibull versus gamma based on likelihood ratio. *Communications in Statistics, Theory and Methods* 19(4):375–381.
- Bayarri, M.J., and J.O. Berger. 1998. Quantifying surprise in the data and model verification. In *Bayesian statistics, 6*, ed. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, 53–82. London: Oxford University Press.
- Berger, J.O., and L.R. Pericchi. 1996. The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association* 91:109–121.
- Berger, J.O., and L.R. Pericchi. 2001. Objective Bayesian methods for model selection: introduction and comparison. In *Model selection*, ed. P. Lahiri. Beachwood: Institute of Mathematical Statistics.
- Cohen, A.C. 1969. A generalization of Weibull distribution. Marshall Space Flight Centre, NASA Contractor report No. 61293, Cont. NAS8-11175.
- Dempster, A.P. 1974. The direct use of likelihood for Significance testing. *Proceedings of Conference on Foundational Questions in Statistical Inference, University of Aarhus*, 335–352.
- Devroye, L. 1986. *Non-uniform random variate generation*. New York: Springer-Verlag.
- Gelfand, A.E., and S.K. Ghosh. 1998. Model choice: A minimum posterior predictive loss approach. *Biometrika* 85:1–11.
- Gelman, A., X.L. Meng, and H.S. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* 6:733–807.
- Laud, P.W., and J.G. Ibrahim. 1995. Predictive model selection. *Journal of the Royal Statistical Society, Series B* 57:247–262.
- Lawless, J.F. 1982. *Statistical Models and Methods for Lifetime Data*. New York: Wiley.
- Lieblein, J., and M. Zelen. 1956. Statistical investigation of the fatigue life of deep-groove ball bearings. *Journal of Research of the National Bureau of Standards* 57:273–316.
- Mann, N.R., R.A. Schafer, and N.D. Singpurwalla. 1974. *Methods for statistical analysis of reliability and lifetime data*. New York: Wiley.
- Martz, H.F., and R.A. Waller. 1982. *Bayesian reliability analysis*. New York: Wiley.
- O'Hagan, A. 1995. Fractional Bayes factors for model comparisons. *Journal of the Royal Statistical Society, Series B* 57:99–138.
- Raiffa, H., and R. Schlaifer. 1961. *Applied statistical decision theory*. Cambridge: Harvard University Press.
- Robert, C.P., and G. Casella. 2004. *Monte Carlo statistical methods*. New York: Springer.
- Rubin, D.B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics* 12:1151–1172.
- Sahu, S.K., and D.K. Dey. 2000. A comparison of frailty and other models for bivariate survival data. *Lifetime Data Analysis* 6:207–228.
- Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- Singpurwalla, N.D. 2006. *Reliability and risk: A Bayesian perspective*. New York: Wiley.
- Smith, A.F.M., and G.O. Roberts. 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* 55:3–25.
- Spiegelhalter, D.J., N.G. Best, B.P. Carlin, and L.A. Vander. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:583–640.
- Stacy, E.W. 1962. A generalization of the gamma distribution. *Annals of Mathematical Statistics* 33:1187–1192.
- Upadhyay, S.K. and Mukherjee, B. 2008. Assessing the value of the threshold parameter in the Weibull distribution using Bayes paradigm. Accepted, *IEEE Transactions on Reliability*.
- Upadhyay, S.K., and M. Peshwani. 2003. Choice between Weibull and lognormal models: A simulation based Bayesian study. *Communications in Statistics. Theory and Methods* 32:381–405.
- Upadhyay, S.K. and Smith, A.F.M. 1993. Bayesian inference in the life testing and reliability via Markov chain Monte Carlo simulation. TR-93-19, Department of Mathematics, Imperial College, London.
- Upadhyay, S.K., and A.F.M. Smith. 1994. Modelling complexities in reliability and the role of simulation in Bayesian computation. *International Journal of Continuing Engineering Education* 4:93–104.
- Upadhyay, S.K., N. Vasistha, and A.F.M. Smith. 2001. Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation. *Sankhya, Series A* 63:15–40.