Loss modeling using Burr mixtures

DOI:

[10.1007/s00181-017-1269-7](https://doi.org/10.1007/s00181-017-1269-7)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Abu Bakar, S. A., Nadarajah, S., & Adzhar, Z. A. A. K. (2017). Loss modeling using Burr mixtures. *Empirical Economics*, 1-14. <https://doi.org/10.1007/s00181-017-1269-7>

Published in:

Empirical Economics

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Loss modeling using Burr mixtures

by

S. A. Abu Bakar¹, Saralees Nadarajah² and Z. A. ABSL Kamarul Adzhar¹

¹Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

²School of Mathematics, University of Manchester, Manchester M13 9PL, UK

Abstract: The first ever real data application of a two-component Burr mixture distribution is provided. It is fitted to three loss data sets: fire loss claims in Denmark, fire loss claims for three building categories in Belgium and fire loss data in Norway. Each of these data sets exhibits significant bimodality. The fits of the two-component Burr mixture distribution are compared to those of five other two-component mixture distributions: the two-component Weibull mixture, two-component gamma mixture, two-component Pareto mixture, two-component lognormal mixture and the two-component exponential mixture distributions. The Burr mixture distribution is shown to give the best fit for each data set. The relative performances of the fitted distributions was assessed in terms of Akaike information criterion values, Bayesian information criterion values, consistent Akaike information criterion values, corrected Akaike information criterion values, Hannan-Quinn criterion values, density plots and probability-probability plots.

Keywords: Loss; Maximum likelihood; Mixture distributions

JEL classification: C1

1 Introduction

Bimodal distributions are new in insurance loss modeling. Some recent examples include: distributions of fire and allied losses exhibit a bimodal resemblance (Vernic et al. 2009); Lee and Lin (2010) reason that property and casualty losses are derived from various sources and thus they often show multimodal characteristics.

Bimodal characteristics are commonly modeled by mixture distributions. Some recent applications of mixture distributions to insurance loss modeling include: estimation of catastrophic losses in the United States using Erlang mixture distributions (Lee and Lin 2010); modeling of motor insurance claims in Thailand using lognormal mixture distributions (Satayatham and Talangtam 2012).

Only the Weibull mixture, gamma mixture, Pareto mixture, lognormal mixture, exponential mixture and beta mixture distributions have featured in the applications involving non-negative continuous data. One important mixture distribution that has not featured in applications is the Burr mixture distribution. The Burr distribution is one of the most versatile distributions in statistics. As shown by Rodriguez (1977) and Tadikamalla (1980), the Burr distribution contains the shape characteristics of the normal, lognormal, gamma, logistic and exponential distributions as well as a significant portion of the Pearson type I, II, V, VII, IX and XII families.

Published papers known to us with respect to the Burr mixture distribution are: Ahmad

(1994, 1995), Ahmad et al. (2011a, 2011b), Ahn et al. (2011), Saleem et al. (2011), Ahmad et al. (2012), Daniyal et al. (2014), Sindhu et al. (2014a, 2014b), Daniyal and Rajab (2015) and Al-Moisheer and Sultan (2016). Of these Ahmad (1994, 1995), Ahmad et al. (2011a, 2011b), Ahmad et al. (2012), Daniyal et al. (2014), Sindhu et al. (2014a, 2014b), Daniyal and Rajab (2015) and Al-Moisheer and Sultan (2016) describe methodological developments only or methodological developments followed by simulation studies. Ahn et al. (2011) describe an application to a data from an ongoing matched case-control study of colorectal cancer, but the Burr mixture is used only as a link function. Saleem et al. (2011) describe an application to failure times of electronic valves, but the model used is a mixture of Pareto distributions, a particular case of the mixture of Burr distributions.

The aim of this note is to provide the first ever data applications of the Burr mixture distribution. The results show that the Burr mixture distribution consistently outperforms the known mixture distributions for the data considered. It is hoped that the results of this note could encourage the Burr mixture distribution being applied to innovative problems.

The contents of this note are organized as follows. Section 2 describes six mixture distributions, including the Burr mixture distribution. An application of these distributions to three fire loss data sets is described in Section 3. The efficiencies of two methods for fitting the distributions are compared in Section 4. Some conclusions and future work are noted in Section 5.

2 Mixture models

A two-component mixture distribution can be specified by the density function

$$f(x) = wg(x; \Theta_1) + (1 - w)g(x; \Theta_2), \quad (1)$$

where $0 < w < 1$ is a mixing parameter, g is the density function of each component and Θ_i is a vector of parameters for the i th component. The n th moment corresponding to (1) can be expressed as

$$E(X^n) = w \int_0^\infty x^n g(x; \Theta_1) dx + (1 - w) \int_0^\infty x^n g(x; \Theta_2) dx.$$

If $x_1, x_2, \dots, x_n$ is a random sample from the two-component mixture distribution the log-likelihood function can be expressed as

$$\ln L(\Theta) = \sum_{i=1}^n \ln \{wg(x_i; \Theta_1) + (1 - w)g(x_i; \Theta_2)\}, \quad (2)$$

where $\Theta = (w, \Theta_1, \Theta_2)$.

We take g to correspond to a three-parameter Burr, two-parameter Weibull, two-parameter gamma, two-parameter Pareto, two-parameter lognormal and one-parameter exponential distributions. We shall refer to the resulting mixture distributions as the Burr mixture, Weibull mixture, gamma mixture, Pareto mixture, lognormal mixture and exponential mixture distributions. The expressions for $g(x; \Theta_i)$ and $\int_0^\infty x^n g(x; \Theta_i) dx$ corresponding to these distributions are given in the appendix.

The six two-component mixture distributions were fitted to the data by the method of maximum likelihood, i.e., by maximizing (2) for g taking the six possible forms given by (3) to (8). The maximization was performed by using the function `optim` in R (R Core Development 2015). A wide range of initial values were chosen for maximization, including

- each component of Θ_i taking $0.01, 0.02, \dots, 10$ for $i = 1, 2$;
- the mixing parameter w taking $0.01, 0.02, \dots, 0.99$;
- the method of moments estimates given by the simultaneous solutions of $E(X^i) = (1/n) \sum_{j=1}^n x_j^i$, $i = 1, 2, \dots, 7$ for the Burr mixture distribution;
- the method of moments estimates given by the simultaneous solutions of $E(X^i) = (1/n) \sum_{j=1}^n x_j^i$, $i = 1, 2, \dots, 5$ for the Weibull, gamma, Pareto and lognormal mixture distributions;
- the method of moments estimates given by the simultaneous solutions of $E(X^i) = (1/n) \sum_{j=1}^n x_j^i$, $i = 1, 2, 3$ for the exponential mixture distribution.

`optim` always converged but the estimate for Θ was not unique. Mixture distributions are known to have likelihood surfaces with multiple local maxima (McLachlan and Peel 2000). We chose the maximum likelihood estimate for Θ as the estimate say $\hat{\Theta}$ for which $\ln L(\hat{\Theta})$ was the largest.

Many of the fitted distributions are not nested. Discrimination among them was performed using various criteria:

- the Akaike information criterion due to Akaike (1974) defined by

$$\text{AIC} = 2k - 2 \ln L(\hat{\Theta}),$$

where $\hat{\Theta}$ denotes the maximum likelihood estimate of Θ and k the number of parameters;

- the Bayesian information criterion due to Schwarz (1978) defined by

$$\text{BIC} = k \ln n - 2 \ln L(\hat{\Theta});$$

- the consistent Akaike information criterion (CAIC) due to Bozdogan (1987) defined by

$$\text{CAIC} = -2 \ln L(\hat{\Theta}) + k(\ln n + 1);$$

- the corrected Akaike information criterion (AICc) due to Hurvich and Tsai (1989) defined by

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n-k-1};$$

- the Hannan-Quinn criterion due to Hannan and Quinn (1979) defined by

$$\text{HQC} = -2 \ln L \left(\hat{\Theta} \right) + 2k \ln \ln n;$$

The smaller the values of these criteria the better the fit. For more discussion on these criteria, see Burnham and Anderson (2004) and Fang (2011).

Finally, the classical likelihood ratio test (Cox and Hinkley 1974) can be used to discriminate among nested models. According to this test, if Model 1 has n_1 parameters and yields a log-likelihood of $\ln L_1$ and Model 2, a sub-model of Model 1, has $n_2 < n_1$ parameters and yields a log-likelihood of $\ln L_2$, then the former should be preferred if $2(\ln L_1 - \ln L_2) > \chi_{n_1 - n_2, 0.95}^2$, where $\chi_{\nu, \alpha}^2$ denotes the 100α percentile of a chisquare random variable with ν degrees of freedom.

3 Data application

Here, we illustrate the performance of the mixture distributions in Section 2 using three real data sets. The data sets chosen are the fire loss data for Denmark, Belgium and Norway. Note that all these data sets are skewed and carry extreme tails. Multimodality is not clear from graphical observations. However, various tests are available in the literature to test for multimodality. We performed Hartigans dip test (Hartigan and Hartigan 1985). The test minimizes the maximum difference between the empirical and theoretical distribution functions. All the three data sets passed the Hartigans dip test for multimodality which indicates that they are at least bimodal.

The first data set consists of 2492 fire loss claims in Denmark recorded by Copenhagen Reinsurance from a period from 1980 to 1990. The data are in millions of Danish Kroner (DKK). They have been properly adjusted to reflect inflation for 1985 values. Thorough discussion of this data has been made by McNeil (1997) and Resnick (1997). The data can be found in Embrechts et al. (2013). The following are some summary statistics for the data: the minimum is 0.3134, the median is 1.634, the mean is 3.063, the maximum is 263.300 and the standard deviation is 7.977.

The second data set was collected by the reinsurance broker AON Re Belgium and consists of 1823 fire loss claims for three building categories. Claim sizes used for this purpose were recorded in thousands of Belgium Franc. Some summary statistics for the data are as follows: the minimum is 0.15, the median is 9.07, the mean is 363.50, the maximum is 190500.00 and the standard deviation is 4868.259.

The final data set used in this study is the Norwegian fire loss data. The data consists of 9181 observations for a period of 21 years starting 1972. However, we consider the three recent years observations only, that is, the data from 1990 to 1992 comprising of 1867 observations. The data are rounded to thousands of Norwegian Kroner. Some summary statistics for the data are as follows: the minimum is 0.500, the median is 1.106, the mean is 1.992, the maximum is 102.400 and the standard deviation is 4.384.

The six mixture models proposed in Section 2 were fitted to each of the three data sets. The method of maximum likelihood was used as described in Section 2. For each of the three data sets, the Burr mixture distribution gave the best fit in terms of the six criteria

stated in Section 2. That is, the Burr mixture distribution gave the smallest values for $-\ln L$, AIC, AICc, BIC, HQC and CAIC.

We now give details of the results for the Denmark, Belgium and Norway data sets. The p -values of Hartigan's dip test for these data sets were 0.212, 0.1966 and 0.9923, respectively. The values of $-\ln L$, AIC, AICc, BIC, HQC and CAIC for the fitted distributions for these data sets are shown in Tables 1 to 3. The density functions for the fitted distributions for these data sets are shown in Figs 1 to 3. The probability plots for these data sets are shown in Figs 4 to 6.

The Burr mixture distribution gives the best fit for each of the three data sets in terms of the six criteria, fitted density plots and probability plots. The second best fit is by the lognormal mixture distribution. Other mixture distributions perform variably for the three data sets. The worst fit for the Danish and Belgian fire losses is by the exponential mixture distribution. The worst fit for the Norwegian fire losses is by the Pareto mixture distribution.

Table 1: The values of $-\ln L$, AIC, AICc, BIC, HQC and CAIC for mixture models fitted to Danish fire loss data.

Mixture models	$-\ln L$	AIC	AICc	BIC	HQC	CAIC
Burr	3786.473	7586.946	7586.991	7627.692	7601.741	7634.692
Weibull	4304.553	8619.106	8619.130	8648.210	8629.674	8670.852
Gamma	4162.031	8334.062	8334.086	8363.166	8344.630	8385.808
Pareto	4966.836	9943.672	9943.696	9972.776	9954.240	9995.418
Lognormal	3955.784	7921.568	7921.592	7950.672	7932.136	7973.314
Exponential	4973.005	9952.010	9952.020	9969.473	9958.351	10007.756

Table 2: The values of $-\ln L$, AIC, AICc, BIC, HQC and CAIC for mixture models fitted to AON Re Belgium fire loss data.

Mixture models	$-\ln L$	AIC	AICc	BIC	HQC	CAIC
Burr	8386.478	16786.956	16787.018	16825.514	16801.180	16832.514
Weibull	8572.077	17154.154	17154.187	17181.695	17164.314	17203.712
Gamma	8671.749	17353.498	17353.531	17381.039	17363.658	17403.056
Pareto	8501.022	17012.044	17012.077	17039.585	17022.204	17061.602
Lognormal	8413.339	16836.678	16836.711	16864.219	16846.838	16886.236
Exponential	8853.842	17717.684	17717.697	17745.225	17719.780	17767.242

Table 3: The values of $-\ln L$, AIC, AICc, BIC, HQC and CAIC for mixture models fitted to Norwegian fire loss data.

Mixture models	$-\ln L$	AIC	AICc	BIC	HQC	CAIC
Burr	2245.589	4505.178	4505.238	4543.903	4519.446	4550.903
Weibull	2537.426	5084.852	5084.884	5112.512	5095.044	5134.577
Gamma	2442.712	4895.424	4895.456	4923.084	4905.616	4945.149
Pareto	3006.171	6022.342	6022.374	6050.002	6032.534	6072.067
Lognormal	2327.221	4664.442	4664.474	4692.102	4674.634	4714.167
Exponential	2938.727	5883.454	5883.467	5900.050	5889.569	5937.179

Figure 1: Fitted probability density functions for the Danish fire loss data: solid black line for mixture of Burr, dashed red line for mixture of gamma, dotted green line for mixture of Pareto, dotdash blue line for mixture of Weibull, longdash cyan line for mixture of lognormal and longdotdash magenta line for mixture of exponential models.

Figure 2: Fitted probability density functions for the AON Re Belgium fire loss data: solid black line for mixture of Burr, dashed red line for mixture of gamma, dotted green line for mixture of Pareto, dotdash blue line for mixture of Weibull, longdash cyan line for mixture of lognormal and longdotdash magenta line for mixture of exponential models.

Figure 3: Fitted probability density functions for the Norwegian fire loss data: solid black line for mixture of Burr, dashed red line for mixture of gamma, dotted green line for mixture of Pareto, dotdash blue line for mixture of Weibull, longdash cyan line for mixture of lognormal and longdotdash magenta line for mixture of exponential models.

Figure 4: Probability-probability plots (P-P plots) of the Danish fire loss data: mixture of Burr (black), mixture of Weibull (red), mixture of gamma (green), mixture of Pareto (blue), mixture of lognormal (cyan) and mixture of exponential (magenta) models.

Figure 5: Probability-probability plots (P-P plots) of the Belgian fire loss data: mixture of Burr (black), mixture of Weibull (red), mixture of gamma (green), mixture of Pareto (blue), mixture of lognormal (cyan) and mixture of exponential (magenta) models.

Figure 6: Probability-probability plots (P-P plots) of the Norwegian fire loss data: mixture of Burr (black), mixture of Weibull (red), mixture of gamma (green), mixture of Pareto (blue), mixture of lognormal (cyan) and mixture of exponential (magenta) models.

The Burr mixture distribution has two more parameters than the other mixture distributions fitted. This may explain why it provides better fits in terms of density and probability plots. However, the Burr mixture distribution also provides better fits in terms of the six criteria, which penalize for the additional parameters.

The use of the likelihood ratio test showed that: i) the Weibull mixture distribution provides a significantly better fit than the exponential mixture distribution for all three data; ii) the gamma mixture distribution also provides a significantly better fit than the exponential mixture distribution for all three data; iii) the Burr mixture distribution also provides a significantly better fit than the Pareto mixture distribution for all three data.

4 A simulation study

Here, we assess the efficiency of the `optim` routine in the R software in computing the maximum likelihood estimates versus using the EM algorithm in computing the maximum likelihood estimates. The efficiency is measured in terms of the central processing unit time to compute the maximum likelihood estimates.

We use a simulation study described by the following scheme:

1. simulate a sample of size n from a two-component Burr mixture distribution;
2. compute the maximum likelihood estimates of the parameters by using the `optim` routine;
3. compute the maximum likelihood estimates of the parameters by using the EM algorithm;
4. repeat steps 1 to 3 ten thousand times;
5. repeat steps 1 to 4 for $n = 15, 16, \dots, 100$.

The parameter values in step 1 were taken to be $(\Theta_1, \Theta_2, w) = (1, 2, 2, 1, 0.5)$, $(1, 2, 5, 1, 0.5)$, $(1, 2, 0.2, 1, 0.5)$, $(0.1, 2, 7, 1, 0.5)$ and $(1, 0.2, 2, 0.1, 0.5)$.

Figure 7: The mean central processing unit time (over the ten thousand samples) versus n .

The mean central processing unit time (over the ten thousand samples) versus n is plotted in Fig 7 for the five parameter choices. We see that the use of the `optim` routine is

more efficient for every n and for each parameter values. The relative efficiency of using the `optim` routine increases with increasing n . The maximum likelihood estimates of (Θ_1, Θ_2, w) obtained using the `optim` routine and the EM algorithm were identical up to the decimal places considered.

In the simulations, the true parameter values were taken to be one of five choices. But the pattern in Fig 7 was the same for a wide range of other parameter values. In particular, the `optim` routine was always more efficient for every n and for every parameter choice and the relative efficiency of using the `optim` routine always increased with increasing n .

5 Conclusions

We have provided the first ever real data application of a two-component Burr mixture distribution. This application has involved data sets on fire loss for Denmark, Belgium and Norway. The Burr mixture distribution was shown to give the best fit for each of these data sets when compared to five other two-component mixture distributions. The best fit was assessed in terms of $-\ln L$, AIC, AICc, BIC, HQC, CAIC, fitted density plots and probability plots.

Most of the known applications of mixture distributions have involved only two-component mixtures. There are only a few applications involving multicomponent mixtures. In particular, we are not aware of any application involving multicomponent Burr mixture distributions. A future work is to seek innovative applications of multicomponent mixture distributions.

We like to mention that the fire loss data for the three countries were chosen only to illustrate the superior fit and the first real data application of the Burr mixture distribution. We do not claim that the Burr mixture distribution will always provide better fits than other mixture distributions. Of course there may be data sets where the Burr mixture distribution does not provide as good a fit as other mixture distributions. But in the data sets considered, the Burr mixture distribution does provide better fits than other mixture distributions. The message of the paper to the readers is the following: consider the Burr mixture distribution as a possible model in the future.

A final comment relates to the fitted densities in Figs 1 to 3. Although the data are multimodal (according to Hartigans dip test), the fitted densities do not show clear multimodality. So, we also fitted the data by using non-mixture distributions, including the Burr, Weibull, gamma, Pareto, lognormal and exponential distributions. Of these, the Burr distribution gave the best fit for each data set according to $-\ln L$, AIC, AICc, BIC, HQC, CAIC, fitted density plots and probability plots. But the fit of the two-component Burr mixture distribution was significantly better than the fit of the Burr distribution for each data set. In particular, the values of AIC, AICc, BIC, HQC and CAIC for the two-component Burr mixture distribution were smaller.

Appendix: Densities and moments of the mixture components

The three-parameter Burr distribution has

$$g(x; \Theta_i) = \frac{\alpha_i \beta_i \lambda_i^{\alpha_i \beta_i} x^{\beta_i - 1}}{(\lambda_i^{\beta_i} + x^{\beta_i})^{\alpha_i + 1}} \quad (3)$$

and

$$\int_0^\infty x^m g(x; \Theta_i) dx = \alpha_i \lambda_i^m B\left(\alpha_i - \frac{m}{\beta_i}, \frac{m}{\beta_i} + 1\right)$$

provided that $m < \alpha_i \beta_i$, where $\Theta_i = (\lambda_i, \alpha_i, \beta_i)$ and $B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt$ denotes the beta function. Here, $\alpha_i > 0, \beta_i > 0$ are known as shape parameters while $\lambda_i > 0$ are known as scale parameters. The two-parameter Pareto distribution is the particular case for $\beta_i = 1$.

The two-parameter Weibull distribution has

$$g(x; \Theta_i) = \frac{\alpha_i x^{\alpha_i - 1}}{\lambda_i^{\alpha_i}} \exp\left[-\left(\frac{x}{\lambda_i}\right)^{\alpha_i}\right] \quad (4)$$

and

$$\int_0^\infty x^m g(x; \Theta_i) dx = \lambda_i^m \Gamma\left(1 + \frac{m}{\alpha_i}\right),$$

where $\Theta_i = (\lambda_i, \alpha_i)$ and $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) dt$ denotes the gamma function. Here, $\alpha_i > 0$ are known as shape parameters while $\lambda_i > 0$ are known as scale parameters. The one-parameter exponential distribution is the particular case for $\alpha_i = 1$.

The two-parameter gamma distribution has

$$g(x; \Theta_i) = \frac{x^{\alpha_i - 1} \exp(-x/\lambda_i)}{\lambda_i^{\alpha_i} \Gamma(\alpha_i)} \quad (5)$$

and

$$\int_0^\infty x^m g(x; \Theta_i) dx = \frac{\lambda_i^m \Gamma(\alpha_i + m)}{\Gamma(\alpha_i)},$$

where $\Theta_i = (\lambda_i, \alpha_i)$. Here, $\alpha_i > 0$ are known as shape parameters while $\lambda_i > 0$ are known as scale parameters. The one-parameter exponential distribution is the particular case for $\alpha_i = 1$.

The two-parameter Pareto distribution has

$$g(x; \Theta_i) = \frac{\alpha_i \lambda_i^{\alpha_i}}{(x + \lambda_i)^{\alpha_i + 1}} \quad (6)$$

and

$$\int_0^\infty x^m g(x; \Theta_i) dx = m! \frac{\lambda_i^m \Gamma(\alpha_i - m)}{\Gamma(\alpha_i)}$$

provided that $m < \alpha_i$, where $\Theta_i = (\lambda_i, \alpha_i)$. Here, $\alpha_i > 0$ are known as shape parameters while $\lambda_i > 0$ are known as scale parameters.

The two-parameter lognormal distribution has

$$g(x; \Theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i x} \exp \left[-\frac{(\ln x - \mu_i)^2}{2\sigma_i^2} \right] \quad (7)$$

and

$$\int_0^\infty x^m g(x; \Theta_i) dx = \exp \left(m\mu_i + \frac{m^2\sigma_i^2}{2} \right),$$

where $\Theta_i = (\mu_i, \sigma_i)$. Here, $-\infty < \mu_i < \infty$ are known as location parameters while $\sigma_i > 0$ are known as scale parameters.

The one-parameter exponential distribution has

$$g(x; \Theta_i) = \lambda_i \exp(-\lambda_i x) \quad (8)$$

and

$$\int_0^\infty x^m g(x; \Theta_i) dx = m! \lambda_i^{-m},$$

where $\Theta_i = (\lambda_i)$. Here, $\lambda_i > 0$ are known as rate parameters.

Acknowledgments

The authors would like to thank the Editor and the referee for careful reading and comments which greatly improved the paper.

References

- [1] Ahmad KE (1994) Small sample results for a nonlinear discriminant function estimated from a mixture of two Burr type XII distributions. Computers and Mathematics with Applications 28 : 13-20.
- [2] Ahmad KE (1995) The efficiency of a nonlinear discriminant function based on unclassified initial samples from a mixture of two Burr type XII populations. Computers and Mathematics with Applications 30 : 1-7.
- [3] Ahmad KE, Jaheen ZF, Mohammed HS (2011a) Finite mixture of Burr type XII distribution and its reciprocal: Properties and applications. Statistical Papers 52 : 835-845.
- [4] Ahmad KE, Jaheen ZF, Mohammed HS (2011b) Bayesian estimation under a mixture of the Burr type XII distribution and its reciprocal. Journal of Statistical Computation and Simulation 81 : 2121-2130.

- [5] Ahmad KE, Jaheen ZF, Mohammed HS (2012) Bayesian prediction based on type-I censored data from a mixture of Burr type XII distribution and its reciprocal. *Statistics* 46 : 787-797.
- [6] Ahn J, Mukherjee B, Gruber SB, Sinha S (2011) Missing exposure data in stereotype regression model: Application to matched case-control study with disease subclassification. *Biometrics* 67 : 546-558.
- [7] Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 : 716-723.
- [8] Al-Moisheer AS, Sultan KS (2016) Estimation of a discriminant function from a mixture of two Burr type III distributions. *Communications in Statistics - Simulation and Computation* 45 : 3760-3775.
- [9] Bozdogan H (1987) Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* 52 : 345-370.
- [10] Burnham KP, Anderson DR (2004) Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research* 33 : 261-304.
- [11] Cox DR, Hinkley DV (1974) *Theoretical statistics*. Chapman and Hall, London.
- [12] Daniyal M, Aleem M, Nawaz T (2014) On some classical properties of doubly truncated mixture of Burr XII and Weibull distributions. *American Journal of Theoretical and Applied Statistics* 3 : 55-59.
- [13] Daniyal M, Rajab M (2015) On some classical properties of the mixture of Burr XII and Lomax distributions. *Journal of Statistics Applications and Probability* 4 : 173-181.
- [14] Embrechts P, Klüppelberg C, Mikosch T (2013) *Modelling extremal events for insurance and finance*. Springer Verlag, New York.
- [15] Fang Y (2011) Asymptotic equivalence between cross-validations and Akaike Information Criteria in mixed-effects models. *Journal of Data Science* 9 : 15-21.
- [16] Hannan EJ, Quinn BG (1979) The determination of the order of an autoregression. *Journal of the Royal Statistical Society B* 41 : 190-195.
- [17] Hartigan JA, Hartigan PM (1985) The dip test of unimodality. *Annals of Statistics* 13 : 70-84.
- [18] Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. *Biometrika* 76 : 297-307.
- [19] Lee SCK, Lin XS (2010) Modeling and evaluating insurance losses via mixtures of Erlang distributions. *North American Actuarial Journal* 14 : 107-130.
- [20] McLachlan GJ, Peel D (2000) *Finite mixture models*. John Wiley and Sons, New York.
- [21] McNeil AJ (1997) Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin* 27 : 117-137.

- [22] R Development Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- [23] Resnick SI (1997) Discussion of the Danish data on large fire insurance losses. *Astin Bulletin* 27 : 139-151.
- [24] Rodriguez RN (1977) A guide to the Burr type XII distributions. *Biometrics* 64 : 129-134.
- [25] Saleem M, Mahmood MH, Kiani M (2011) Properties and comparison of the Bayes estimates of the Burr mixture parameters. *Science International (Lahore)* 23 : 1-5.
- [26] Sattayatham P, Talangtam T (2012) Fitting of finite mixture distributions to motor insurance claims. *Journal of Mathematics and Statistics* 8 : 49-56.
- [27] Schwarz GE (1978) Estimating the dimension of a model. *Annals of Statistics* 6 : 461-464.
- [28] Sindhu TN, Aslam M, Feroze N (2014a) Statistical inference for mixed Burr type II distribution using a Bayesian framework. *International Journal of Statistics and Economics* 13.
- [29] Sindhu TN, Feroze N, Aslam M (2014b) Preference of prior for Bayesian analysis of the mixed Burr type X distribution under type I censored samples. *Pakistan Journal of Statistics and Operation Research* 10.
- [30] Tadikamalla PR (1980) A look at the Burr and related distributions. *International Statistical Review* 48 : 337-344.
- [31] Vernic R, Teodorescu S, Pelican E (2009) Two lognormal models for real data. *Annals of Ovidius University, Series Mathematics* 17 : 263-277.