

2018

Modeling Insurance Claims Using Flexible Skewed and Mixture Probability Distributions

Aaron J. Leinwander

University of Wisconsin Eau Claire, aaronleinwander@gmail.com

Mohammad A. Aziz

University of Wisconsin Eau Claire

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Leinwander, A. J., & Aziz, M. A. (2018). Modeling Insurance Claims Using Flexible Skewed and Mixture Probability Distributions. *Journal of Modern Applied Statistical Methods*, 17(1), eP2467. doi: 10.22237/jmasm/1525133100

Modeling Insurance Claims Using Flexible Skewed and Mixture Probability Distributions

Cover Page Footnote

ACKNOWLEDGEMENTS We want to thank Office of Research and Sponsored Programs (ORSP) and Mathematics Department at University of Wisconsin-Eau Claire for supporting our work. We also want to thank New Hampshire Comprehensive Health Care Information Systems (NHCHIS) for providing us the data.

EMERGING SCHOLARS

Modeling Insurance Claims Using Flexible Skewed and Mixture Probability Distributions

Aaron J. Leinwander

University of Wisconsin - Eau Claire
Eau Claire, WI

Mohammad A. Aziz

University of Wisconsin - Eau Claire
Eau Claire, WI

The normal distribution comes as a first choice when fitting real data, but it may not be suitable if the assumed distribution deviates from normality. Flexible skewed distributions are capable of including skewness and taking into account multimodality. They may be applied to find appropriate distributions for describing the claim amounts in insurance. The objective is to model insurance claims using a set of flexible skewed and mixture probability distributions, and to test how well they fit the claims. Results indicate the skew- t distribution and alpha-skew Laplace distribution are able to describe unimodal claims accurately, whereas scale mixture of skew-normal and skew- t distributions are better alternatives to both unimodal and bimodal conventional distributions such as skew-normal, alpha skew-normal, and mixture of normal distributions. The tail risk measures such as value at risk and tail value at risk are estimated as judgment criteria to assess the fitness of the models.

Keywords: Flexible skewed distributions, skew-normal distribution, skew- t distribution, scale mixture of skew-normal distribution, value at risk, tail value at risk

Introduction

The normal distribution is widely used for a variety of applications in statistics and many other fields because of its simplicity. Insurance data often have non-normal distributions because they are highly skewed and have heavy tails. They may also be bimodal. The skew-normal and skew- t distributions were discussed by Azzalini (1985) and can model the skewness and kurtosis of the data very well, but they lack the ability to fit bimodal data. Eling (2012) used two popular data

sets in insurance and showed that the skew-normal and the skew-Student t distributions are reasonably competitive compared to some models when describing insurance data. Bolance et al. (2008) provided evidence that the skew-normal and log-skew-normal distributions can be used to model bivariate claims illustrating data from the Spanish motor insurance industry. Ahn et al. (2012) used the log-phase-type distribution as a parametric alternative in fitting heavy tailed data. Kazemi and Noorzadeh (2015) used two popular data sets in insurance fields to compare the performance of skew-logistic and skew-normal distribution for fitting insurance claims. Extending the work of Eling (2012), the use of more flexible families of distributions is proposed in modeling insurance claims.

Methodology

A set of skewed and flexible skewed distributions were applied to fit unimodal and bimodal insurance claims. Several mixture models were considered for comparison. Model parameters were estimated using maximum likelihood estimation technique, and used AIC and BIC for model comparison. Data were simulated from the fitted distributions to calculate value at risk and tail value at risk, and assessed the performance of different models. All calculations were done using statistical software R.

Unimodal Skewed Distributions

The skew symmetric family of distributions—in particular, the skew normal and skew- t distributions—have received increasing attention in recent years.

Skew-normal distribution

A random variable is said to follow a skew-normal distribution if its probability density function is defined as

$$f(x) = 2\phi(x)\Phi(\alpha x), x \in R \quad (1)$$

where ϕ represents the standard normal density function and Φ represents the distribution function of a standard normal density and α is a real number that regulates the shape of the distribution.

To fit real data, work with an affine transformation $Y = \mu + \alpha X$ with $\mu \in R$ and $\sigma > 0$. The density of Y then becomes

$$f(y) = \frac{2}{a} f\left(\frac{y-m}{s}\right) F\left(a\frac{y-m}{s}\right), y \in R \quad (2)$$

Denote this density as $Y \sim SN(\mu, \sigma, \alpha)$ which reduces to standard skew-normal distribution in (1) where $Y \sim SN(0, 1, \alpha)$. If α is set to 0 both in (1) and (2), the distributions become the pdf of a standard normal distribution. The Skew-normal densities are plotted in Figure 1 for some selected values of alpha.

Skew-t distribution

The skew-t distribution is similar in nature to the skew-normal distribution. It is an appropriate model that allows us to regulate both skewness and kurtosis of a distribution.

Let Z be a standard skew-normal random variable and W be a variable with $\chi^2(v)$ distribution. Then, if Z and W are independent,

$$X = \frac{Z}{\sqrt{\frac{W}{v}}} \quad (3)$$

the linear transformation $Y = \mu + \sigma X$ has a skew-t distribution with parameters μ, σ and α and we introduce the notation $ST(\mu, \sigma, \alpha)$ to denote the skew-t random variable Y .

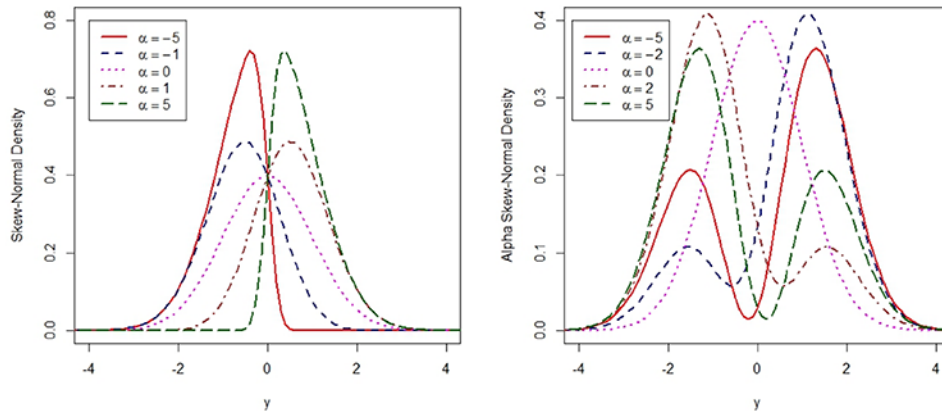


Figure 1. SN and ASN densities for some selected values of the shape parameter

Bimodal Skewed Distributions

Skew normal and skew- t distributions are typically used to fit unimodal data. Recently some other models surfaced that are not only skewed but also capable of modelling multimodality of the data.

Alpha-skew-normal distribution

Elal-Olivero (2010) introduced a new class of skew-normal distributions called alpha-skew-normal distributions, which are skewed and can fit a bimodal data.

A continuous random variable X has an alpha-skew-normal distribution with a probability density

$$f(x) = \frac{(1 - \alpha x)^2 + 1}{2 + \alpha^2} \phi(x), x \in R \quad (4)$$

where α represents the shape parameter.

Denote this density as $X \sim ZSN(\mu, \sigma, \alpha)$. Adjusting the pdf to include location and scale parameters the density becomes

$$f(y) = \frac{\left[1 - a \left(\frac{y - m}{s}\right)\right]^2 + 1}{2 + a^2} f\left(\frac{y - m}{s}\right), y \in R \quad (5)$$

If α is set to 0 in 2.4 and 2.5, the standard normal distribution is obtained. The graph of the alpha-skew-normal distribution for some selected values of the skewness parameter α is shown in Figure 1.

Alpha-Skew-Laplace Distribution

Alpha-skew-Laplace distribution (ASLP) is introduced in Harandi & Alamatsaz (2013), which is an extension of Elal-Olivero's alpha-skew distribution that uses Laplace distribution instead of normal distribution. This distribution can be both unimodal and bimodal with the suitable choices of the values of its parameters.

A continuous random variable X is said to follow a Alpha-Skew-Laplace distribution if its pdf has the form

$$f(x) = \frac{(1-\alpha x)^2 + 1}{4(1+\alpha^2)} e^{-x}, x \in R \quad (6)$$

where α represents the shape parameter. Alpha-Skew-Laplace random variable is denoted by ASLP(α).

Suppose $X \sim \text{ASLP}(\alpha)$. Then, ASLP density of location and scale is defined as the distribution of $Y = \mu + \sigma X$ for $\mu \in R$ and $\sigma > 0$. The corresponding density function is given by

$$f(x) = \frac{\left(1 - a \frac{y-m}{s}\right)^2 + 1}{4s(1+a^2)} e^{-\frac{|y-m|}{s}}, x \in R \quad (7)$$

where $\theta = (\mu, \sigma, \alpha)$.

Mixture Models

Consider several conventional mixture distributions and recently developed scale mixture of skew-normal and skew- t distributions.

In the mixture model context the density of x is expressed as a mixture of P parametric densities such that

$$f(y) = \sum_i^p \pi_i f(x; \theta_i) \quad (8)$$

where $\pi_i \geq 0$, $i = 1, 2, \dots, p$ with $\sum_i^p \pi_i = 1$, are called mixing weights of the p^{th} component of the mixture, which is characterized by parameter θ_i and $\psi = (\pi_1, \pi_2, \dots, \pi_{(p-1)}, \theta_1, \theta_2, \dots, \theta_p)$ denotes the vector of parameters of the model.

Two components Weibull Mixture

The density function of mixture of a Weibull (WW) distribution is given by

$$g(x; a_1, s_1, a_2, s_2, w) = w.f(x; a_1, s_1) + (1-w)f(x; a_2, s_2) \quad (9)$$

where α is the shape parameter and σ is the scale parameter with Weibull density

$$f(x; \sigma, \alpha) = \frac{\alpha}{\sigma} \left(\frac{x}{\sigma} \right)^{\alpha-1} e^{-\left(\frac{x}{\sigma}\right)^\alpha}, \quad x \geq 0 \quad (10)$$

Equation (9) can be modified to include the location parameter.

Mixture Weibull and lognormal Distribution

The probability density function of the mixture of Weibull and lognormal density (LNWL) is written as

$$g(x; S, a, q, l, w) = w \cdot f(x; S, a) + (1 - w) f(x; q, l) \quad (11)$$

with Weibull density given in (10) and lognormal density given by

$$f(x; \theta, \lambda) = \frac{1}{x\theta\sqrt{2\pi}} e^{-\frac{(\ln(x)-\lambda)^2}{2\theta^2}}, \quad -\infty < x < \infty \quad (12)$$

Finite Mixture of Scale Mixture of Skew Normal Distribution

Suppose $Z \sim SN(0, \sigma^2, \alpha)$ and U be a positive random variable, independent of Z , with distribution function $H(u; \nu)$. Then the random variable $Y = \mu + U^{-1}Z$, where $\mu \in R$ is a location parameter, is said to follow a scale mixture of skew normal (SMSN) distribution if its pdf is given by

$$f(y) = \int_0^\infty f(y; m, S^2, u^{i-1}) F\left(u^{\frac{1}{2}} a \left(\frac{y - m}{S}\right)\right) dH(u) \quad (13)$$

In the definition $H(\cdot; \nu)$ is known as the mixing scale distribution, for each choice of this we get different members of the family like normal, skew-normal and student- t . A finite mixture of SMSN distributions [8,9] is a density defined as in (8) where the i^{th} component of the mixture is a SMSN density with parameters μ_i , σ_i^2 , α_i , and ν_i . For simplicity we assume $\nu_1 = \nu_2 = \dots = \nu$.

Estimation Method

Estimate the parameters of all models using Maximum likelihood estimation technique. For all of the models considered, in the first step the log likelihood

function is written using the probability density functions provided. In the second step, the GenSA function from R package ‘GenSA’ is used for optimization.

Risk Measures

A risk measure provides the information contained in the distribution function of a random variable in one single real number. Risk measures are useful to evaluate and monitor the risk exposures of investors. Most commonly used risk measures in the field of insurance and finance are the p -quantile risk measures, based on a percentile concept. Consider value at risk (VaR) and tail value at risk (TVaR) for illustration. For other risk measures refer to Dhaene et. al (2006).

Value at risk

The value at risk at level p , is the amount of capital required to ensure that the enterprise does not become technically insolvent. In probabilistic terms, the VaR at level p is defined as the $100p\%$ quantile of the distribution of the terminal wealth. More precisely, for any $p \in (0,1)$, the p -quantile measure or VaR for a random variable X , denoted by $Q_p[X]$, is defined as

$$Q_p[X] = \inf \{x \in R \mid F_X(x) \geq p\}$$

Tail value at risk

As with the VaR, the tail value at risk or TVaR is defined using some confidence level. The TVaR is intuitively defined as the expected value of the loss, given the loss is greater than the VaR. TVaR is the arithmetic average of the VaR’s of the loss. TVaR at confidence level p , given the p -quantile risk measure Q_p , is defined as

$$TVaR_p(X) = E(X \mid X > Q_p(x)).$$

Results

New Hampshire Dental Claims

The data, a subset of 4,849 claims, were obtained from New Hampshire Comprehensive Health Care Information Systems (NHCHIS) from 2013 (NHCHIS, 2013). Both the original and log of the data are analyzed. The

histograms of claims and log of claims are shown in Figure 2. As observed from the histograms, the distribution of claims is clearly right skewed with at least one peak. The log of the data normalizes the set however; the second mode is magnified in the data, which may provide interesting results while modeling with flexible skewed distributions. The normal probability plots for original data and log data are shown in Figure 3. In addition to the normal plot, perform Shapiro-Wilk normality test for the original and transformed data and in both cases small p-values ($< 2.2e-16$) reject the normality of the data.

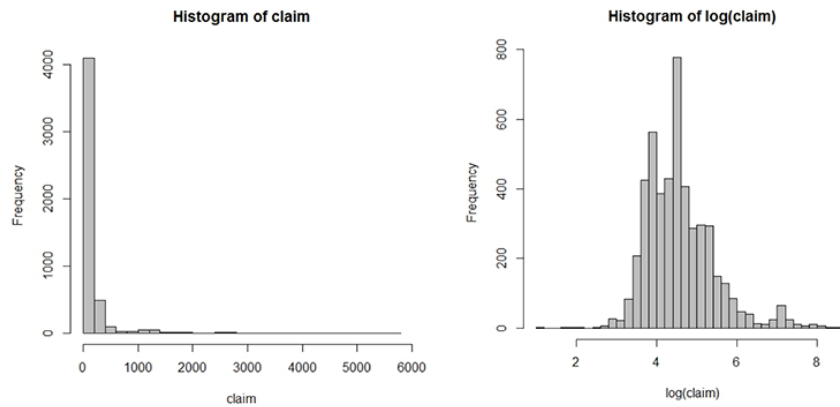


Figure 2. Histograms of claims and log(claims)

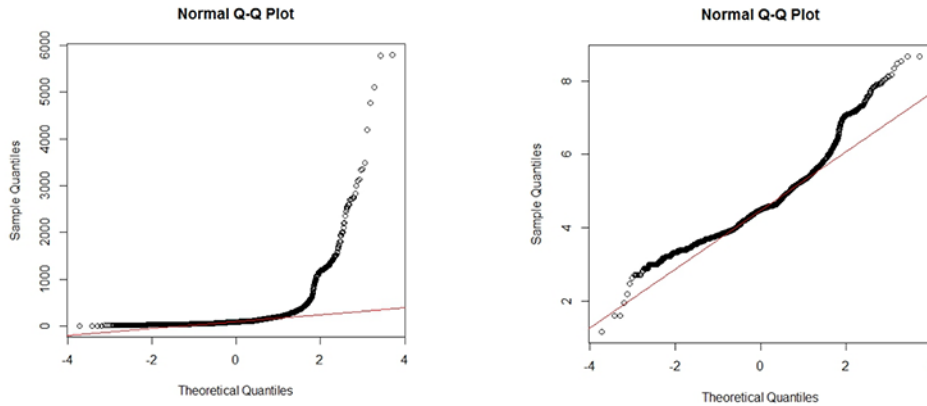


Figure 3. Normal Q-Q plot of claims and log(claims)

MODELING CLAIMS WITH FLEXIBLE PROBABILITY DISTRIBUTIONS

Model Summary

The model fitting criteria Akaike Information Criterion (AIC) and Bayesian Information criteria (BIC) are presented in Table 1, and the parameter estimates for different models are presented in Table 2.

Table 1. Model Fitting Summary

Flexible Distributions	AIC		BIC	
	Claim	Log(Claim)	Claim	Log(Claim)
$LN(\mu, \sigma)$	56,490.57	11,577.01	56,503.55	11,589.98
$SN(\mu, \sigma, \alpha)$	63,975.42	11,495.62	63,986.39	11,506.59
$ST(\mu, \sigma, \alpha, \nu)$	55,628.93	11,319.93	55,637.91	11,328.90
$ASN(\mu, \sigma, \alpha)$	67,790.68	11,522.68	67,801.65	11,533.65
$ASLP(\mu, \sigma, \alpha)$	60,839.63	11,789.47	60,850.60	11,800.44

Table 2. Estimated parameter values for Flexible models

Model	Parameters	Claim	Log(Claim)
LN	μ	4.57052140	1.50355520
	σ	0.84797760	0.17743440
SN	μ	14.40608000	3.61953500
	σ	352.31400000	1.27414300
	α	67.34349000	3.12904000
ST	μ	29.77575800	3.70813000
	σ	62.62235700	1.02769000
	α	9.88083700	2.60815900
	ν	1.40193600	6.98577800
ASN	μ	443.16673600	5.47591400
	σ	286.24893900	0.91448270
	α	1.75006100	1.10426600
ASLP	μ	66.00000000	4.44265093
	σ	93.33526570	0.61567602
	α	-0.32263370	-0.07233172

From the model fitting criteria while considering original claim amount, we observe that skew- t distribution is dominating all other distributions having a very small AIC and BIC values followed by lognormal distribution.

The observed and expected densities for different skewed distributions for original claim amount are presented in Figure 4. The figure depicts the same pattern as observed from the AIC and BIC values. In addition, ASLP model seems to fit the data quite well too which was the third best model in terms of the AIC and BIC values. When the transformed or log data are considered, skew- t

distribution is still the best model followed by skew-normal distribution and alpha-skew-normal distribution.

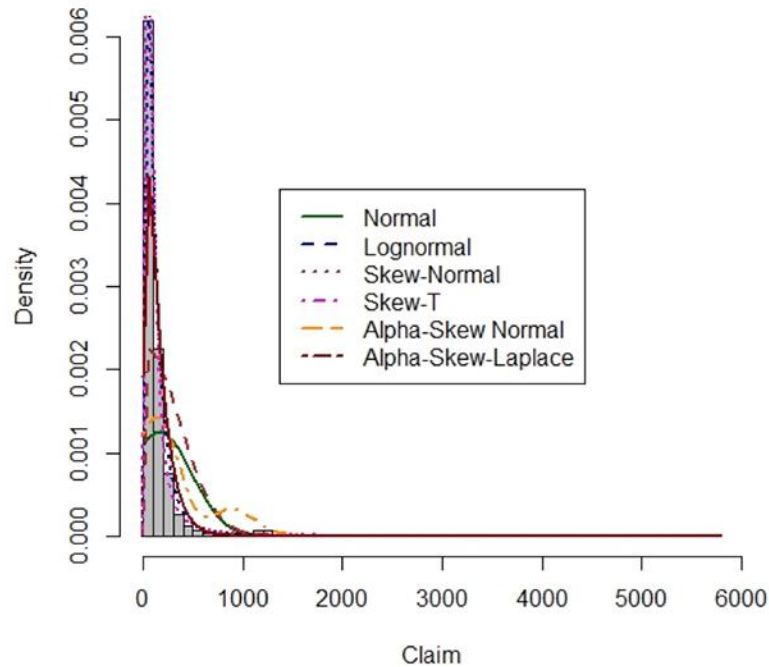


Figure 4. Observed and expected densities for claims

The observed and expected densities for different skewed distributions for $\log(\text{claims})$ are shown in Figure 5. From the fitted densities in Figure 5, we observed that alpha-skew-normal distribution not only fits the data well it is also capable of taking into account the second mode of the distribution. For the mixture models AIC and BIC, values are presented in Table 3.

MODELING CLAIMS WITH FLEXIBLE PROBABILITY DISTRIBUTIONS

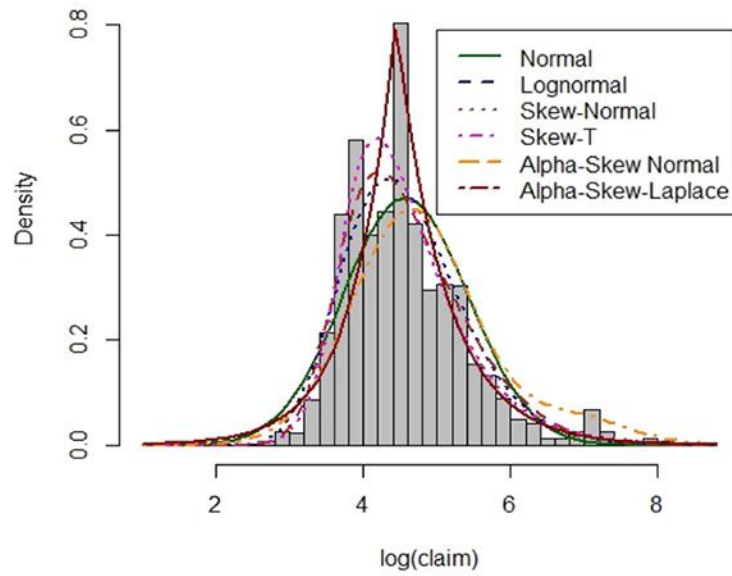


Figure 5. Observed and expected densities for transformed data

Table 3. Model Fitting Summary for mixture distributions

Mixture Distributions	AIC		BIC	
	Claim	Log(Claim)	Claim	Log(Claim)
WW	56,265.65	11,528.59	56,268.65	11,531.56
NN	58,096.27	11,338.31	58,128.70	11,370.74
TT	57,206.87	11,329.87	57,239.30	11,362.30
SMSN	55,994.35	11,356.33	56,039.76	11,401.73
SMST	55,501.11	11,354.49	55,546.52	11,399.89
LNWL	55,852.66	11,409.20	55,957.63	11,416.18

From the model fitting criteria we observed that for the original claim amount scale mixture of skew- t (SMST) distributions best fit the data followed by SMSN. The lognormal and Weibull mixture (LNWL) is the third best model.

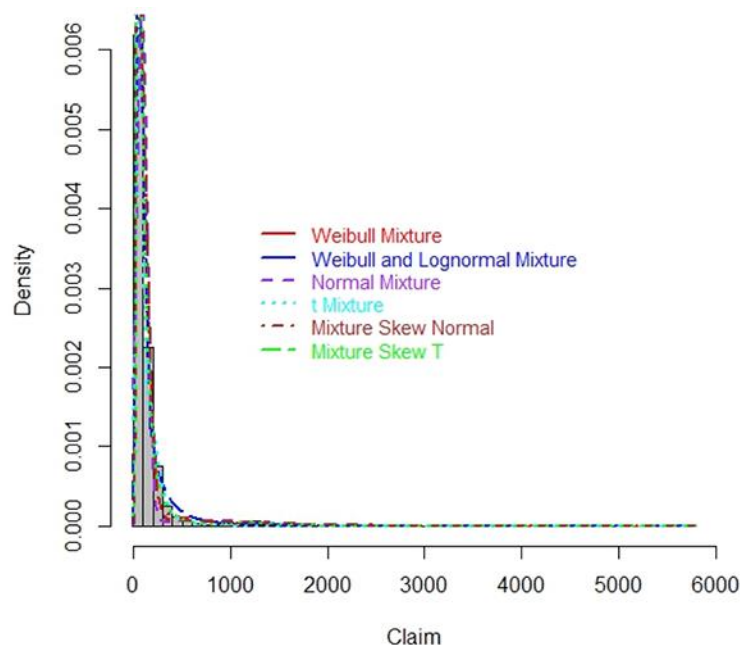


Figure 6(a). Observed and expected densities for mixture distributions using original and transformed data

The observed and expected densities for different mixture models for claim and $\log(\text{claims})$ are shown in [Figure 6a](#) and [6b](#). Most of the skewed distributions fit the original claim amount quite well. For the $\log(\text{claims})$, besides skew- t and skew-normal, mixture of lognormal and Weibull distribution shows better fit. Unfortunately, none of the two-component fitted models was able to show the second model of the distribution. A modality test ([Xu et al., 2014](#)) was performed to find the actual number of modes for the transformed data. The modality test shows that there are at least three modes for the transformed data. Three and four components mixture models were then fitted to the data and the observed and expected densities for the models are plotted in [Figure 7](#).

MODELING CLAIMS WITH FLEXIBLE PROBABILITY DISTRIBUTIONS

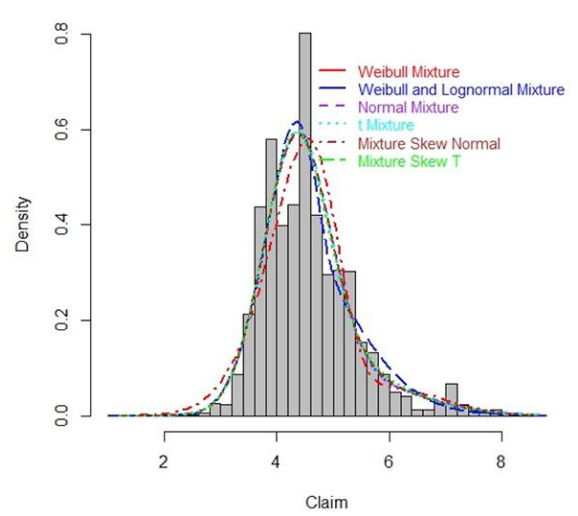


Figure 6(b). Observed and expected densities for mixture distributions using transformed data.

From the two and three component mixtures we observe that either a three-component skew t -mixture or four-component t -mixture model will fit the transform data quite well.

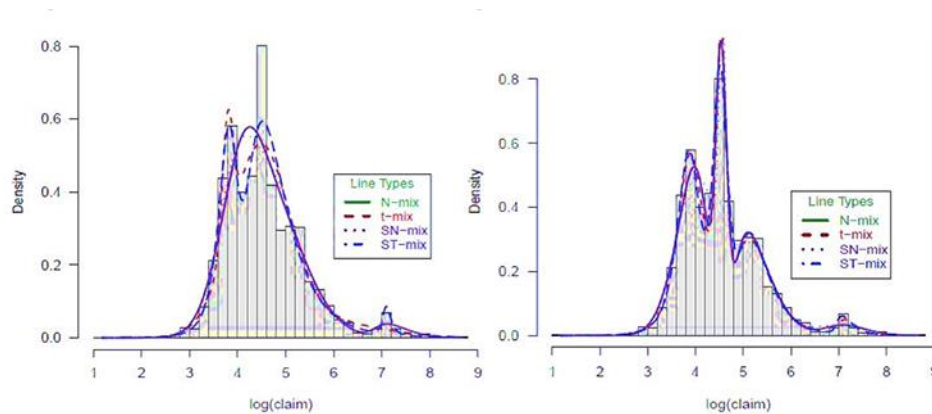


Figure 7. Observed and expected densities for transformed data with 3 and 4 component mixtures respectively

Simulation and Estimation of risk measures

A million observations were simulated from the assumed model using estimated parameters for both skewed and mixture models. Then, find estimators for VaR and TVaR for each of the model and compare them with the empirical values. Next, plot the risk measurement results for varying confidence levels (theoretical formulas for VaR and TVaR were not derived for the models considered). The VaR and TVaR for the skewed models for the transformed data are shown Figure 8 and for the mixture models are presented in Figure 9.

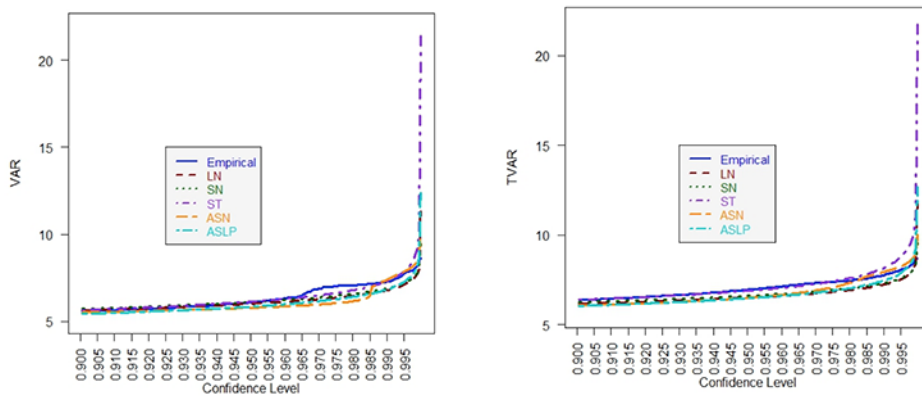


Figure 8. VaR and TVaR for the skewed models for the original data

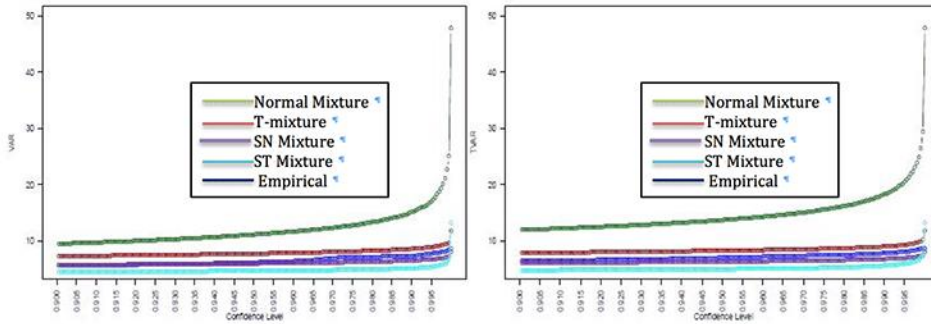


Figure 9. VaR and TVaR for the mixture models for the original data

MODELING CLAIMS WITH FLEXIBLE PROBABILITY DISTRIBUTIONS

From the figures, we observe that for original data both VaR and TVaR values are very close to the empirical values with the TVaR values very stable for almost all distributions. While mixture models are considered, all of the models provided very close estimates as that of the empirical ones except mixture of normal.

Conclusion

A comparison of skewed, flexible and mixture distributions for modelling insurance claims were compared using AIC, BIC and density estimation. Conventional distributions were examined, as well as constructed flexible skew distributions and their mixtures. Many insurance claims have a relative skewness to them, and in some cases there can be two modes to the data. The results indicated skew- t distribution seems consistently to be one of the top choices in modeling unimodal insurance data. However, this unimodal skewed distribution fails to account for the second mode of the data. In the case of the transformed dental claims, it is arguable that the alpha-skew distributions have advantages in modeling because it was able to show the second mode of the distribution. It is clear that the skew distributions will work well when modeling single mode with skewed data. These distributions were created in order to model this type of data, but in case of bimodal data, flexible distributions will have the advantage. The more bimodal a data set appears, the better the overall fit will be from flexible distributions compared to the others.

References

- Ahn, S., Kim, J. H. T., & Ramaswami, V. (2012). A new class of models for heavy tailed distributions in finance and insurance. *Insurance: Mathematics and Economics*, 51(1), 43–52. doi: 10.1016/j.insmatheco.2012.02.002
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171-178.
- Bolance, C., Guillen, M., Pelican, E., & Vernic, R. (2008). Skewed bivariate models and non-parametric estimation for the CTE risk measure. *Insurance: Mathematics and Economics*, 43(3), 386-393. doi: 10.1016/j.insmatheco.2008.07.005
- Dhaene, J., Vanduffel, S., Tang, Q., Goovaerts, M., Kaas, R., & Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic Models*, 22(4), 573-606. doi: 10.1080/15326340600878016

- Elal-Olivero, D. (2010). Alpha-skew-normal-distribution. *Proyecciones: Journal of Mathematics*, 29(3), 224-240. doi: 10.4067/s0716-09172010000300006
- Eling, M. (2012). Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? *Insurance: Mathematics and Economics*, 51(2), 239-248. doi: 10.1016/j.insmatheco.2012.04.001
- Ferreira, C. D. S., Bolfarine, H., & Lachos, V. H. (2011). Skew scale mixtures of normal distributions: Properties and estimation. *Statistical Methodology*, 8(2), 154-171. doi: 10.1016/j.stamet.2010.09.001
- Harandi, S. S., & Alamatsaz, M. H. (2013). Alpha-skew-laplace distribution. *Statistics and Probability Letters*, 83(3), 774-782. doi: 10.1016/j.spl.2012.11.024
- Kazemi, R., & Noorizadeh, M. (2015). A comparison between skew-logistic and skew-normal distributions. *Matematika*, 31(1), 15-24. Retrieved from <http://www.matematika.utm.my/index.php/matematika/article/download/742/679>
- New Hampshire Comprehensive Health Care Information System. (2013). [Medical and dental claims]. Unpublished raw data.
- Prates, M. O., Cabral, C. B., & Lachos, V. H. (2013). mixsmsn: fitting finite mixture of scale mixture of skew-normal distributions. *Journal of Statistical Software*, 54(12), 1-20. doi: 10.18637/jss.v054.i12
- Xu, L., Bedrick, E. J., Hanson, T., & Restrepo, C. (2014). A comparison of statistical tools for identifying modality in body mass distributions. *Journal of Data Science*, 12, 175-196.