Czesław Domański[*], *Katarzyna Bolonek-Lasoń*[**]

GENERALIZATIONS OF TUKEY-LAMBDA DISTRIBUTIONS

Abstract. The Generalized Lambda Distribution (GλD) is a four-parameter generalization of Tukey's Lambda family. Several methods for estimating the parameters of the GλD have been reported in the literature, but the most popular is the moment-method matching proposed by Ramberg and Schmeiser (1974). One criticism of the GλD referred to above is that the shape parameters also determine skewness. It seems reasonable that there should be three linear parameters determining position, scale, and skewness and two parameters determining the shapes of the two tails. This suggests a natural generalization of the GλD to give a five-parameter lambda distribution (FPLD). The aim of paper is to show that the GλD and the FPLD describe empirical distribution quite well.

Key words: generalized lambda distribution, estimation methods, distribution fitting, quotations.

I. THE GENERALIZED λ DISTRIBUTION

One of the more important tasks in statistical data analysis is fitting a probability distribution. We usually search for distribution parameters through fitting a probability density function (PDF) or a cumulative distribution function (CDF) to some data. In the case of Tukey-lambda distribution we search for distribution parameters on the basis of the percentile function which is the inverse of the cumulative distribution function.

An example of the percentile function is Tukey's λ function

$$R(p) = \frac{p^\lambda - (1-p)^\lambda}{\lambda}, \qquad 0 \le p \le 1, \quad \lambda \ne 0. \tag{1}$$

This function depends on one parameter, so it describes only a family of symmetrical continuous probability distributions. For example,

- $\lambda = -1$ approximately Cauchy
- $\lambda = 0$ exactly logistic

---

[*] Professor, Department of Statistical Methods, University of Lodz.
[**] Ph.D., Department of Statistical Methods, University of Lodz.

- $\lambda = 0.14$ approximately normal
- $\lambda = 0.5$ U-shaped
- $\lambda = 1$ exactly uniform $(x \in (-1,1))$.

Ramberg and Schmeiser (1974) generalized Tukey's λ distribution to a four-parameter distribution defined by the percentile function

$$R(p) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2}, \qquad 0 \le p \le 1, \tag{2}$$

where λ_1 is a location parameter, λ_2 is a scale parameter, λ_3 is a skewness parameter and λ_4 is a kurtosis one.

The percentile function is the inverse of the cumulative distribution function, so if X is a continuous random variable with percentile function R and p is a uniform random variable on the interval $\langle 0,1 \rangle$, then $X = R(p)$ and CDF of variable X has form

$$F(R(p)) = p. \tag{3}$$

Differentiating both sides of the equation (3) with respect to p yields the expression for density function

$$f(x) = \frac{1}{\dfrac{dR(p)}{dp}} = \frac{\lambda_2}{\lambda_3 p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}}, \qquad 0 \le p \le 1, \tag{4}$$

which depends only on three parameters. Unfortunately, the CDF does not exist in analytic form.

The generalized λ distribution defines the family of asymmetrical probability distributions of continuous variable, inter alia: Gamma distribution, Beta or Weibull distribution. Besides, the most important thing is that, if we search for parameters of the distribution, we do not have to know which theoretical distribution will better describe the empirical data.

The density function of GλD is not proper function for all combinations of values of shape parameters. Ramberg and Schmeiser (1974) indicated four regions of parameters which produce a proper statistical distribution (Table 1); however Karian, Dudewicz and McDonald (1996) characterized two new regions. Figure 1 shows all allowed regions of λ_3 and λ_4 parameter space.

Table 1. Range of λ values for Ramberg's parametrization of the GλD

Region	distribution parameters				Minimum	Maximum
	λ_1	λ_2	λ_3	λ_4		
1	all	< 0	< -1	> 1	$-\infty$	$\lambda_1 + (1/\lambda_2)$
2	all	< 0	> 1	< -1	$\lambda_1 - (1/\lambda_2)$	∞
3	all	> 0	> 0	> 0	$\lambda_1 - (1/\lambda_2)$	$\lambda_1 + (1/\lambda_2)$
	all	> 0	$= 0$	> 0	λ_1	$\lambda_1 + (1/\lambda_2)$
	all	> 0	> 0	$= 0$	$\lambda_1 - (1/\lambda_2)$	λ_1
4	all	< 0	< 0	< 0	$-\infty$	∞
	all	< 0	$= 0$	< 0	λ_1	∞
	all	< 0	< 0	$= 0$	$-\infty$	λ_1

Source: King R.A.R., MacGillivray H.L., A Starship estimation method for Generalized λ distributions, Austral. & New Zealand J. Statist. 41(3), 1999, 353-374.
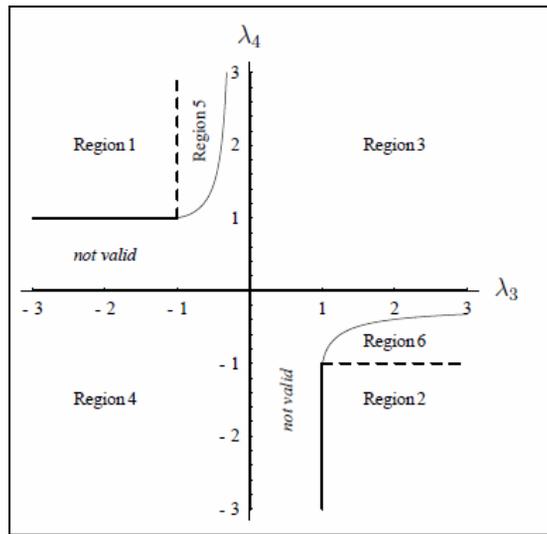
Fig. 1. Regions of the λ_3 and λ_4 plane that produce proper statistical distributions
in the GλD

II. METHODS FOR ESTIMATING THE PARAMETERS OF THE GλD

Several methods for estimating the parameters of the GλD have been reported in the literature (Lakhany and Mausser (2000), Taristano (2010), Taristano (2004), King and MacGillivray (1999), Karvanen and Nuutinen (2007), Su (2007)); we describe briefly some of them.

1) The moment-matching method was proposed by Ramberg and Schmeiser (1979).

This method can be described as follows. First, we find the expressions for the mean, the variance, and the third and fourth moments of GλD:

$$\mu = \lambda_1 + \frac{1}{\lambda_2}\left(\frac{1}{\lambda_3+1} - \frac{1}{\lambda_4+1}\right)$$

$$\sigma^2 = \frac{1}{\lambda_2^2}\left(A_2 - A_1^2\right)$$

$$\beta_3 = \frac{1}{\left(A_2 - A_1^2\right)^{\frac{3}{2}}}\left(A_3 - 3A_1A_2 + 2A_1^3\right)$$ (5)

$$\beta_4 = \frac{1}{\left(A_2 - A_1^2\right)^2}\left(A_4 - 4A_1A_3 + 6A_1^2A_2 + 3A_1^4\right)$$

where

$$A_1 = \frac{1}{\lambda_3+1} - \frac{1}{\lambda_4+1}$$

$$A_2 = \frac{1}{2\lambda_3+1} + \frac{1}{2\lambda_4+1} - 2\beta(\lambda_3+1,\lambda_4+1)$$

$$A_3 = \frac{1}{3\lambda_3+1} - \frac{1}{3\lambda_4+1} - 3\beta(2\lambda_3+1,\lambda_4+1) + 3\beta(\lambda_3+1,2\lambda_4+1)$$

$$A_4 = \frac{1}{4\lambda_3+1} + \frac{1}{4\lambda_4+1} - 4\beta(3\lambda_3+1,\lambda_4+1) + 6\beta(2\lambda_3+1,2\lambda_4+1) - 4\beta(\lambda_3+1,3\lambda_4+1)$$

Next we equate the mean, the variance, and the third and fourth moments of GλD to the corresponding mean μ^*, variance $(\sigma^*)^2$, skewness β_3^*, and kurtosis β_4^* of the sample. Finally, we compute the parameters λ from equations

$$\mu = 0$$
$$\sigma^2 = 1$$
$$\beta_3 = \beta_3^*$$ (6)
$$\beta_4 = \beta_4^*$$

Ramberg at all (1979) presented the tables of the parameters λ for selected values of β_3^*, β_4^* and $\mu = 0$, $\sigma^2 = 1$.

2) The least squares method was proposed by Öztürk and Dale (1985). This method can be described as follows. Let x_i, $i = 1,...,n$ denote the ith order statistic of data which is to be represented by the quantile function $R(u)$ and let U_i, $i = 1,...,n$ denote the order statistic of the corresponding uniformly distributed random variable. The method finds values of λ for which the differences between the observed and predicted order statistics are as small as possible. So, we must minimize the function

$$G(\lambda) = \sum_{i=1}^{n} \left(x_{(i)} - \lambda_1 - \frac{Z_i}{\lambda_2} \right)^2 \tag{7}$$

where:

$$Z_i = \frac{1}{\lambda_3} \left(EU_{(i)}^{\lambda_3} - 1 \right) - \frac{1}{\lambda_4} \left(E\left(1 - U_{(i)}\right)^{\lambda_4} - 1 \right)$$

$$EU_{(i)}^{\lambda_3} = \frac{\Gamma(n+1)\Gamma(i+\lambda_3)}{\Gamma(i)\Gamma(n+\lambda_3+1)}$$

$$E\left(1 - U_{(i)}\right)^{\lambda_4} = \frac{\Gamma(n+1)\Gamma(n-i+\lambda_4+1)}{\Gamma(n-i+1)\Gamma(n+\lambda_4+1)}$$

3) The starship method can be described in a few steps:

Step 1. We select a region in four-dimensional space that covers the range of the four parameters λ.

Step 2. On the region selected in Step 1, we overlay a four-dimensional rectangular grid.

Step 3. We evaluate the grid points created in Step 3 by performing a goodness-of-fit test on the corresponding distributions. If the test is satisfied, we stop procedure; otherwise we continue with the next point in the gird.

Many articles concerning the GλD are devoted to the comparison of methods of estimating parameters. Authors show that the moment matching method and the least square method fit probability distribution into empirical data better then the starship method. For that reason, we use only the first two estimate methods in our calculations.

III. APPLICATIONS

In this section we focus our attention on the application of GλD to stock market quotations. The first data concern daily quotations of WIG20 index from the time period 03.01.1994-26.10.2012. Using the moment matching and the least square methods we calculate the parameters of GλD. The first four moments of the data are:

$$\mu^* = 1901,203 \qquad \beta_3^* = 0,6$$
$$\sigma^* = 733,404 \qquad \beta_4^* = 2,74 \qquad (8)$$

In moment matching method for the values of skewness and kurtosis, mentioned above, we are inside the range of tabulated values in paper Ramberg at all (1979), so the lambda values were read directly from the table (Ramberg at all (1979)). However, in the least square method we found the lambda values from equation (7). Table 2 includes the results of both methods.

Table 2. The estimate parameters of GλD for daily quotations of WIG20 index

Parameter	Estimate method	
	Moment matching	Least squares
λ_1	1188,34	1177,24
λ_2	0,0003	0,0003
λ_3	0,0355	0,0346
λ_4	0,3265	0,3357

Source: own calculations.
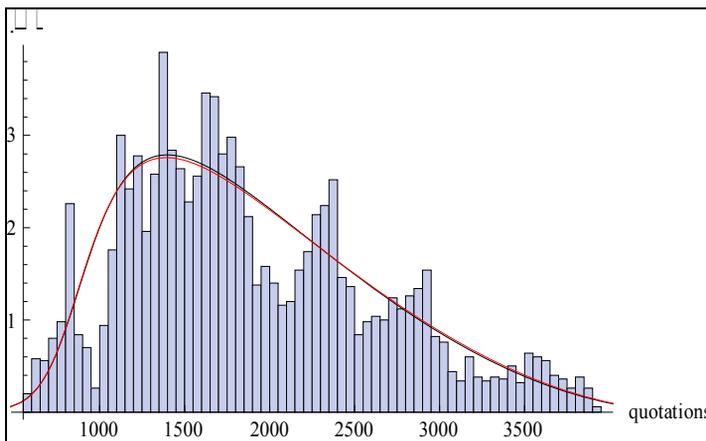
Fig. 2. The histogram of daily quotations of WIG20 index and corresponding GλD
(red curve – the solution for the moment matching method,
black curve – the solution for the least squares method)

Figure 2 illustrates the results of fitting a GλD to the quotation distribution of WIG20 index. The difference between the two GλDs is almost imperceptible. Ramberg at all (1979) calculated χ^2 goodness-of-fit statistics to check that the model fits the data well. In our cases the computed value of χ^2 is large, so we must reject the solutions for the quotation of WIG20 index. Of course, we should expect these results. Besides, these data form a very large set, so in the next example we chose a smaller set of data than the first.

Now we consider daily quotations of Nikkei 225 index from the time period 18.08.2003-29.07.2005. The first four moments of the data read:

$$\mu^* = 11135,54 \qquad \beta_3^* = -0,26$$
$$\sigma^* = 486,97 \qquad \beta_4^* = 2,47 \tag{9}$$

We found the distribution parameters of quotations of Nikkei 225 index in the same way as the WIG20 index. Table 3 shows the results of fitting GλD.

Table 3. The estimate parameters of GλD for daily quotations of Nikkei 225 index

Parameter	Estimate method	
	Moment matching	Least squares
λ_1	10750,8	10889,8
λ_2	0,00058	0,00084
λ_3	0,0843	0,2343
λ_4	0,4294	0,6544

Source: own calculations.

In this case the goodness of fit test rejected both solutions but the computed values of χ^2 were smaller than WIG20 case. This result may suggest that GλD does not describe well the asymmetrical distributions. It could be caused by the fact that the shape parameter also determine skewness, so maybe there should be three linear parameters determining position, scale, and skewness and two parameters determining the shapes of the two tails. This suggests a natural generalization of the GλD to give a five-parameter lambda distribution (FPLD).

The five-parameter version of generalized lambda distribution was proposed by Gilchrist (2000). The quantile function is given by

$$Q(\lambda) = \lambda_1 + \frac{\lambda_2}{2}\left((1-\lambda_3)\frac{p^{\lambda_4}-1}{\lambda_4} - (1+\lambda_3)\frac{(1-p)^{\lambda_5}-1}{\lambda_5}\right) \tag{9}$$

where $0 \le p \le 1$, $\lambda_2 \ge 0$, $\lambda_3 \in \langle -1,1 \rangle$. According to the method proposed by Öztürk and Dale (1985), we must minimize the function

$$G(\lambda) = \sum_{i=1}^{n} \left(x_{(i)} - Z_{(i)} \right)^2 \tag{10}$$

where

$$Z_{(i)} = \lambda_1 + \frac{(1-\lambda_3)\lambda_2}{2\lambda_4} \left(\frac{\Gamma(n+1)\Gamma(i+\lambda_4)}{\Gamma(i)\Gamma(n+1+\lambda_4)} - 1 \right) +$$

$$+ \frac{(1+\lambda_3)\lambda_2}{2\lambda_5} \left(1 - \frac{\Gamma(n+1)\Gamma(n+1-i+\lambda_5)}{\Gamma(n+1-i)\Gamma(n+1+\lambda_5)} \right)$$

The lambda values for the five-parameter lambda distribution are:

$$\begin{matrix} \lambda_1 = 11144{,}1 & \lambda_2 = 810{,}001 & \lambda_3 = 0{,}7 \\ \lambda_4 = 0{,}17789 & \lambda_5 = 0{,}400532 & \end{matrix} \tag{11}$$

The computed value of χ^2 equals 17,09 for which p-value is equal 0,1952. So, this time the model fits the data quite well.
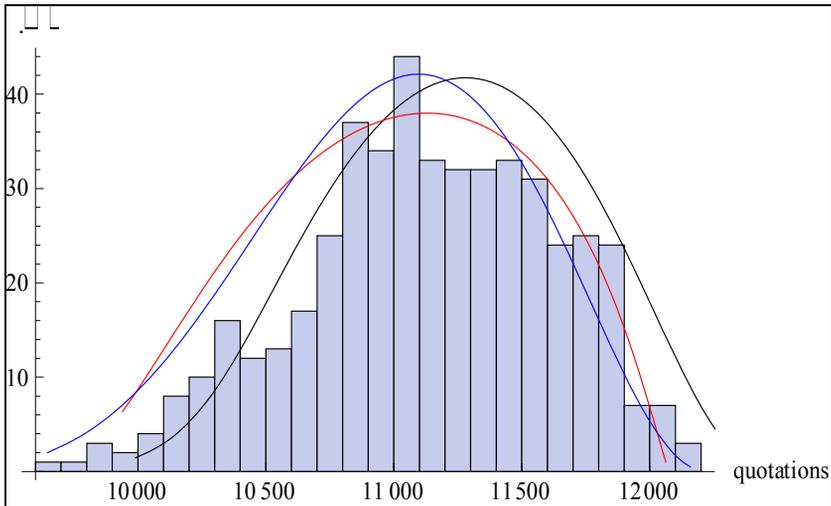
Fig. 3. The histogram of daily quotations of Nikkei 225 index and corresponding GλD and FPLD (black curve – the GλD fitting by the moment matching method, red curve – the GλD fitting by the least squares method, blue curve – the FPLD)

IV. CONCLUSION

We considered the four- and five-parameter lambda distributions. We used two methods of estimating the distribution parameters: moment matching and least square ones. The empirical analysis shows that FLDP describes the stock market quotations better than GλD. Of course, the distribution found describe quotations only for a given period time. We do not claim that daily quotations of WIG20 index or Nikkei 225 index have probability distribution with parameters shown in Table 2 and 3. It is sometimes claimed in the literature that generalizations of Tukey-lambda distributions are not easy to use because it is difficult to find the distribution parameters. However, in this paper we show that GλD and FLDP may be used to describe any asymmetric distribution.

REFERENCES

Domański Cz., Pruska K. (2000) Nieklasyczne metody statystyczne, Polskie Wydawnictwo Ekonomiczne, Warszawa.

King R.A.R., MacGillivray H.L., A starship estimation method for the generalized λ distributions, Austral. & New Zealand J. Statist. 41(3), 1999,353-374.

Lakhany A., Mausser H., Estimating the parameters of generalized lambda distribution, Algo Research Quarterly, vol. 3, No.3, December 2000.

Ramberg J. S. and Schmeister B. W. (1974). An approximate method for generating asymmetric random variables. Canon. ACM, 17, 78-82.

Ramberg J. S., Tadikamalla P.R., Dudewicz E.J., Mykytka E.F., (1979), A probability distribution and its uses in Fitting Data, Technometries 21 nr 2, s. 201-214.

Su S., Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R, Journal of Statistical Software, October 2007, Volume 21, Issue 9.

Tarsitano A., Fitting the generalized lambda distribution to income data, COMPSTAT 2004 Symposium, Physica-Verlag/Springer 2004, 1861-1867.

Tarsitano A., Comparing estimation methods for the FPLD, Journal of Probability and Statistics, Volume 2010, Article ID 295042.

Czesław Domański, *Katarzyna Bolonek-Lasoń*

UOGÓLNIENIA ROZKŁADÓW LAMBDA-TUKEYA

Uogólniony rozkład lambda (GλD) jest czteroparametrowym uogólnieniem rodziny rozkładów lambda-Tukeya. W literaturze możemy odnaleźć wiele metod estymacji parametrów GλD, jednak najbardziej popularną metodą jest metoda momentów zaproponowana przez Ramberga i Schmeisera (1974). Wskazuje się, że jedną z wad uogólnionego rozkładu lambda jest to, że parametry kształtu określają również skośność rozkładu. Wydaje się, że powinny być trzy parametry określające położenie, skalę i skośność oraz dwa parametry określające kształt ogonów rozkładu. Z tego względu dokonano uogólnienia GλD poprzez wprowadzenie pięcio-parametrowego rozkładu lambda (FPLD).