*STATISTICA, anno LXXVI, n. 4, 2016*

# THE GENERALIZED DOUBLE LOMAX DISTRIBUTION WITH APPLICATIONS

Abu Seif Mohammad-Fares[1]
*Higher Institute For Administrative Development, Damascus University, Halboni, Damascus, Syria.*
Vajjha Venkata HaraGopal
*Department of Statistics, Science College, Osmania University, Hyderabad, India.*

## 1. Introduction

The heavy-tailed distributions have recently received a lot of attention in the literature, their role in shaping the base of extreme value and robustness theories made them so attractive, also their appearance in several fields like finance, communications and social sciences made them essential in dealing with arising problems involving extreme events such as value at risk in finance.
In this paper we propose a new heavy tailed distribution which is a generalization of the double Lomax distribution introduced by Bindu (2011) who applied it on IQ data, the proposed distribution was characterized by Freimer and Mudholkar (1989), the generalized double Pareto density coming from the same family of the proposed distribution was used by Armagan and Dunson (2011) as a heavy tailed prior in the Bayesian approach, we have derived some general properties of the generalized double Lomax distribution and then illustrated its usefulness empirically by fitting it to financial data from the American stock market and comparing it with the most widely used distributions.
The remaining of this paper is organized as follows: The generalized double Lomax distribution (GDL) is introduced in the second section along with its cumulative distribution, quantile, some general properties and MLE of its parameters. A simulation study is conducted in the third section. GDL is fitted for Several data sets of daily returns from the American stock market in the fourth section and we conclude in the fifth section.

## 2. The GDL Distribution

A continuous random variable is said to follow Generalized Double Lomax distribution (GDL) if its probability density function is given by:

$$f(x) = \frac{v}{2s(1+\frac{|x-m|}{s})^{v+1}} \quad ; \quad s,v>0 \; , \; x,m \in R \qquad (1)$$

[1] Corresponding Author. E-mail: fares26285@hotmail.com

*2.1. CDF*

The Cumulative distribution function is given by:

$$F(x) = \begin{cases} \frac{1}{2\left(1-\dfrac{x-m}{s}\right)^{v}} & \text{if} \quad x < m \\ \frac{1}{2} & \text{if} \quad x = m \\ 1 - \frac{1}{2\left(1+\dfrac{x-m}{s}\right)^{v}} & \text{if} \quad x > m \end{cases}$$

*2.2. Quantiles*

The quantile function is given by:

$$X = \begin{cases} m + s\left[1 - e^{\frac{-\ln(2F)}{v}}\right] & \text{if} \quad F \in [0, \frac{1}{2}] \\ m + s\left[e^{\frac{-\ln(2(1-F))}{v}} - 1\right] & \text{if} \quad F \in [\frac{1}{2}, 1] \end{cases}$$

*2.3. Central moments*

The moments are given by:

$$\mathbf{E}(x-m)^r = \int_{-\infty}^{\infty} (x-m)^r f(x)dx = \frac{s^r \Gamma(r+1)\Gamma(v-r)[-1^r+1]}{2\Gamma(v)} \quad ; \quad r \in Z \quad v > r \quad (2)$$

*2.4. Central tendency measures*

The mean (with $v > 1$), median and mode are all same and equal to $m$

*2.5. Variance*

The variance is given by:

$$Var[x] = \frac{s^2 2\Gamma(v-2)}{\Gamma(v)} \quad ; \quad v > 2$$

*2.6. Mean absolute deviations*

The mean absolute deviation around both the mean and median is given by:

$$\mathbf{E}(|x-m|) = \int_{-\infty}^{\infty} |x-m| f(x; m, s, v) d(x) = \frac{s}{v-1} \qquad v > 1$$

## 2.7. Skewness and kurtosis

Since all odd moments are zeros the skewness is also zero (with $v > 3$). The excess kurtosis is given by:

$$k = \frac{\mu_4}{{\mu_2}^2} - 3 = \frac{6\Gamma(v)\Gamma(v-4)}{\Gamma^2(v-2)} - 3 \quad ; \quad v > 4$$

## 2.8. Entropy

Shannons entropy is defined as follows:

$$H(x) \quad = \quad -\int_{-\infty}^{\infty} \ln[f(x)]f(x)dx = \ln\left(\frac{2se^{(\frac{v+1}{v})}}{v}\right) \tag{3}$$

## 2.9. Fisher Information

Let $\theta = (\theta_1, \theta_2, \theta_3) = (m, s, v)$ then the $(i,j)^{th}$ element of Fisher information matrix is:

$I(\theta)_{i,j} = -\mathbf{E}[\frac{d^2}{\theta_i\theta_j}\ln(f(x,\theta))|\theta], \qquad i = 1..3 \quad j = 1..3$

And the Fisher information matrix is given by:

$$I(\theta) = \begin{pmatrix} \frac{1}{v^2} & \frac{-1}{s(v+1)} & 0 \\ \frac{-1}{s(v+1)} & \frac{v}{s^2(v+2)} & 0 \\ 0 & 0 & \frac{v(v+1)^2}{s^2(v+2)} \end{pmatrix}$$

## 2.10. Maximum Likelihood Estimation

Let $(x_1, x_2, ...., x_n)$ be a random sample which follows the non-standardized generalized double Lomax distribution then the likelihood function is given by:

$$L = L(x,m,s,v) \quad = \quad \prod_{i=1}^{n} f_i(x_i,m,s,v) = \frac{v^n}{2^n s^n \prod_{i=1}^{n}(1+\frac{|x_i-m|}{s})^{v+1}}$$

$$LL = \ln(L) = n\ln(v) - n\ln(s) - n\ln(2) + (-v-1)\sum_{i=1}^{n}\ln(1+\frac{|x_i-m|}{s})$$

$$\begin{aligned} \frac{dLL}{dv} \quad &= \quad 0 \Rightarrow \frac{n}{v} - \sum_{i=1}^{n}\ln(1+\frac{|x_i-m|}{s}) = 0 \\ v \quad &= \quad \frac{n}{\sum_{i=1}^{n}\ln(1+\frac{|x_i-m|}{s})} \end{aligned} \tag{4}$$

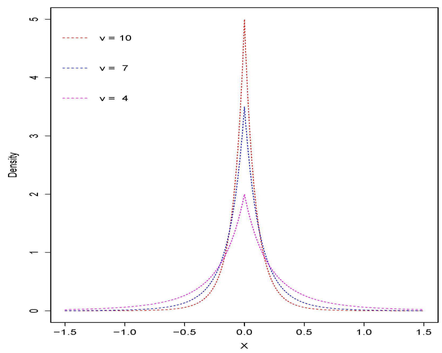


Figure 1 –: Three GDL densities with identical location and scale parameters (m and s) but different shape parameters v.

$$\frac{dLL}{ds} = 0 \Rightarrow -\frac{n}{s} - (v+1)\sum_{i=1}^{n} \frac{-|x_i - m|}{s^2(1 + \frac{|x_i - m|}{s})} = 0 \tag{5}$$

$$\frac{dLL}{dm} = 0 \Rightarrow \sum_{i=1}^{n} \frac{sign(x_i - m)}{(1 + \frac{|x_i - m|}{s})} = 0 \tag{6}$$

The last three equations should be solved numerically to obtain values of $m$, $s$, and $v$ such that the $LL$ is maximum.

## 3. Simulation

The monte carlo experiment has been performed with 500 iterations and repeated for different sample sizes from 250 to 50000 to assess the convergence of the MLEs to the true parameters of (GDL), We consider the values (0,5,10) to be the true parameters (m,s,v) in the random numbers generating process of inversion method, The simulation results are displayed in Table 1, the consistency of the MLE is well pronounced as the bias and standard deviations decrease with larger sample sizes.

TABLE 1
MLE for several simulated data

| **Sample size** | **MLE and Standard Deviation** | | |
|---|---|---|---|
| **n** | $\hat{m}$ | $\hat{s}$ | $\hat{v}$ |
| 250 | 0.0002 | 11.5688 | 22.2102 |
| | (0.0327) | (15.3233) | (28.2523) |
| 500 | -0.0007 | 8.7381 | 16.8797 |
| | (0.0217) | (11.2936) | (20.7081) |
| 1000 | 0.0000 | 6.5309 | 12.8236 |
| | (0.0160) | (5.3941) | (9.8137) |
| 2000 | -0.0005 | 5.5800 | 11.0725 |
| | (0.0115) | (2.3751) | (4.3874) |
| 4000 | -0.0001 | 5.2417 | 10.4311 |
| | (0.0083) | (1.1860) | (2.1453) |
| 6000 | -0.0001 | 5.1291 | 10.2285 |
| | (0.0065) | (0.8920) | (1.6173) |
| 8000 | 0.0003 | 5.1296 | 10.2399 |
| | (0.0056) | (0.7285) | (1.3308) |
| 10000 | -0.0002 | 5.0935 | 10.1661 |
| | (0.0051) | (0.6952) | (1.2704) |
| 15000 | 0.0001 | 5.0544 | 10.1058 |
| | (0.0040) | (0.5083) | (0.9307) |
| 25000 | -0.0001 | 5.0188 | 10.0350 |
| | (0.0031) | (0.3829) | (0.7007) |
| 50000 | 0.0000 | 5.0109 | 10.0215 |
| | (0.0022) | (0.2655) | (0.4833) |

## 4. Applications

The GDL is fitted to several data sets in this section, the data used in the study consists of daily returns between November 2006 and 2016 for three stock indexes from USA (SP500), (Dow 30) and (Nasdaq), and also daily returns with maximum available period for six equities (SYNT), (MFIN), (HIIQ), (ALBO), (ATEC) and (SIRI). The maximum likelihood estimation has been applied for fitting (GDL) preceded by the same for t-student, Generalized error (GED) and Generalized Hyperbolic Distributions (GHYP), Tables 2, 3, 4 and 5 displays the MLE results of the mentioned data.

It can be seen from Table 2 that all returns data exhibits heavy tails since the shape parameter v of t-student is very low (less than 3).

The estimated values of the shape parameter $\beta$ of the generalized error distribution in Table 3 which is less than 1 for all returns data also indicate the heavy-tailedness of the given data.

The Generalized Hyperbolic distribution of 5 parameters has been fitted (Table 4) to account for both skewness and kurtosis present in the data with location parameter $\mu$, scale $\sigma$, skewness $\gamma$ and two shape parameters $\overline{\alpha}$ and $\lambda$.

Table 5 shows the estimated parameters of GDL distribution obtained by MLE, the values of shape parameter v range between 5 and 10 which are low as expected since the excess kurtosis of GDL is infinite at v=4 and it approaches to 3 as v goes to infinity.

Finally, Table 6 represents comparisons between the above four fitted models in terms of their Akaike and Bayesian information criteria (AIC and BIC), from this table the AIC selects GDL for all data sets except for SP500 and Nasdaq, whereas the values of BIC are minimum for all the nine variables selecting GDL as the best model for fitting the studied data.

TABLE 2
MLE for t-student distribution

| **Index /Equity** | **n** | $\hat{m}$ | $\hat{s}$ | $\hat{v}$ |
|---|---|---|---|---|
| **S&P500** | 2516 | 0.0008 | 0.0070 | 2.3313 |
| **Dow30** | 2516 | 0.0006 | 0.0069 | 2.7247 |
| **Nasdaq** | 2516 | 0.0009 | 0.0084 | 2.7663 |
| **SYNT** | 4853 | 0.0000 | 0.0166 | 2.2386 |
| **MFIN** | 5149 | -0.0003 | 0.0150 | 2.3876 |
| **HIIQ** | 947 | -0.0018 | 0.0209 | 2.4583 |
| **ALBO** | 2389 | -0.0019 | 0.0273 | 2.7657 |
| **ATEC** | 2624 | -0.0010 | 0.0263 | 2.5178 |
| **SIRI** | 2516 | -0.0002 | 0.5608 | 2.0015 |

TABLE 3
MLE for Generalized error distribution

| **Index /Equity** | **n** | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\beta}$ |
|---|---|---|---|---|
| **S&P500** | 2516 | 0.0006 | 0.0127 | 0.8099 |
| **Dow30** | 2516 | 0.0005 | 0.0118 | 0.8027 |
| **Nasdaq** | 2516 | 0.0010 | 0.0137 | 0.9074 |
| **SYNT** | 4853 | 0.0000 | 0.0338 | 0.7135 |
| **MFIN** | 5149 | 0.0000 | 0.0263 | 0.8431 |
| **HIIQ** | 947 | -0.0016 | 0.0364 | 0.8302 |
| **ALBO** | 2389 | 0.0000 | 0.0409 | 0.7186 |
| **ATEC** | 2624 | 0.0000 | 0.0426 | 0.8172 |
| **SIRI** | 2516 | 0.0000 | 0.0252 | 0.9149 |

TABLE 4
MLE for Generalized Hyperbolic distribution

| **Index /Equity** | **n** | $\hat{\lambda}$ | $\hat{\bar{\alpha}}$ | $\hat{\mu}$ | $\hat{\sigma}$ | $\hat{\gamma}$ |
|---|---|---|---|---|---|---|
| **S&P500** | 2516 | -0.25458 | 0.30666 | 0.00114 | 0.01294 | -0.00081 |
| **Dow30** | 2516 | -0.06510 | 0.31030 | 0.00092 | 0.01191 | -0.00068 |
| **Nasdaq** | 2516 | 0.06764 | 0.37950 | 0.00139 | 0.01402 | -0.00091 |
| **SYNT** | 4853 | -0.57003 | 0.26337 | -0.00060 | 0.03247 | 0.00159 |
| **MFIN** | 5149 | -0.58681 | 0.29687 | -0.00080 | 0.02787 | 0.00117 |
| **HIIQ** | 947 | 0.21687 | 0.25847 | -0.00302 | 0.03671 | 0.00310 |
| **ALBO** | 2389 | -0.95168 | 0.33099 | -0.00323 | 0.04638 | 0.00243 |
| **ATEC** | 2624 | -0.96679 | 0.23401 | -0.00144 | 0.04835 | 0.00123 |
| **SIRI** | 2516 | -0.7134 | 0.09702 | -0.00041 | 0.0403 | 0.0015 |

TABLE 5
MLE for Generalized double Lomax distribution

| **Index /Equity** | **n** | $\hat{m}$ | $\hat{s}$ | $\hat{v}$ |
|---|---|---|---|---|
| **S&P500** | 2516 | 0.0007 | 0.0530 | 7.2627 |
| **Dow30** | 2516 | 0.0006 | 0.0500 | 7.3322 |
| **Nasdaq** | 2516 | 0.0010 | 0.0824 | 9.6887 |
| **SYNT** | 4853 | 0.0000 | 0.1167 | 6.7033 |
| **MFIN** | 5149 | 0.0000 | 0.1037 | 6.7835 |
| **HIIQ** | 947 | -0.0015 | 0.1911 | 8.8233 |
| **ALBO** | 2389 | -0.0004 | 0.1750 | 6.6390 |
| **ATEC** | 2624 | 0.0000 | 0.1500 | 5.8446 |
| **SIRI** | 2516 | 0.0000 | 0.0804 | 4.9851 |

TABLE 6
Comparisons between the fitted distributions

| Index /Equity | Information Criterion | t-Student | GED | GHYP | GDL |
|---|---|---|---|---|---|
| **S&P500** | **AIC** | -15514.05 | -15522.46 | -15542.66 | -15538.08 |
| | **BIC** | -15496.56 | -15504.97 | -15513.5 | -15520.59 |
| **Dow30** | **AIC** | -15863.72 | -15888.55 | -15897.2 | -15900.86 |
| | **BIC** | -15846.23 | -15871.06 | -15868.04 | -15883.37 |
| **Nasdaq** | **AIC** | -14943.96 | -14948.17 | -14966.96 | -14955.06 |
| | **BIC** | -14926.47 | -14930.67 | -14937.8 | -14937.57 |
| **SYNT** | **AIC** | -21364.04 | -21355.46 | -21395.12 | -21397.98 |
| | **BIC** | -21344.58 | -21336 | -21362.68 | -21378.52 |
| **MFIN** | **AIC** | -24041.16 | -24059.04 | -24065.7 | -24089.06 |
| | **BIC** | -24021.52 | -24039.4 | -24032.97 | -24069.42 |
| **HIIQ** | **AIC** | -3811.748 | -3822.544 | -3823.3 | -3826.806 |
| | **BIC** | -3797.188 | -3807.984 | -3799.034 | -3812.246 |
| **ALBO** | **AIC** | -8574.862 | -8553.606 | -8577.218 | -8587.539 |
| | **BIC** | -8557.526 | -8536.27 | -8548.325 | -8570.203 |
| **ATEC** | **AIC** | -9410.516 | -9435.268 | -9412.692 | -9443.606 |
| | **BIC** | -9392.899 | -9417.651 | -9383.33 | -9425.989 |
| **SIRI** | **AIC** | -10960.99 | -10887.74 | -10973.02 | -10988.24 |
| | **BIC** | -10943.50 | -10870.25 | -10943.87 | -10970.74 |

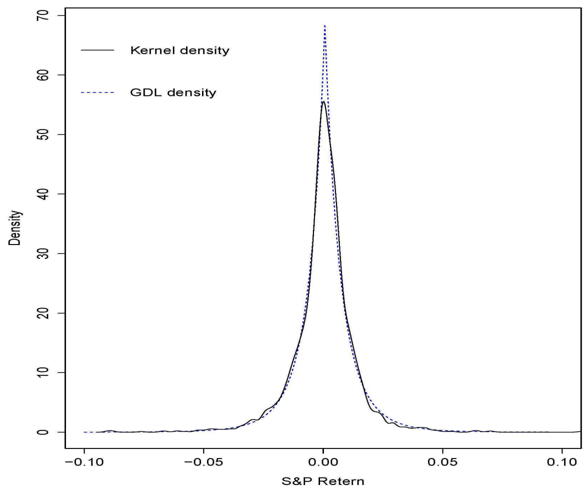


Figure 2 –: Kernel and fitted GDL densities of SP500.
Estimated Kernel density (solid) of the daily returns on the SP 500 index compared with GDL density using MLE (dashed) .

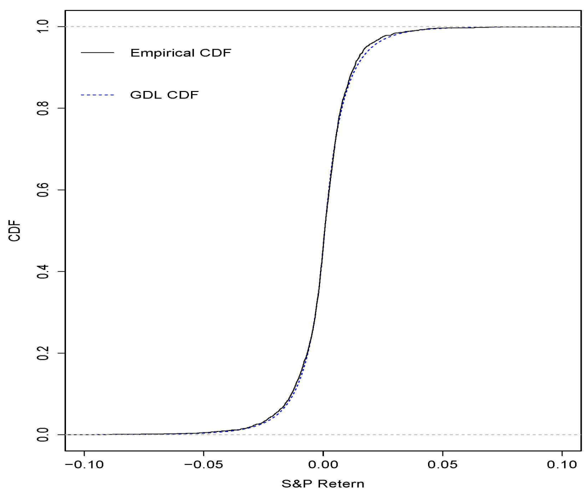


Figure 3 –: CDF's of Empirical and fitted GDL of SP500.
Empirical cumulative distribution function (solid) of the daily returns on the SP 500 index compared with GDL cumulative distribution function using MLE (dashed)

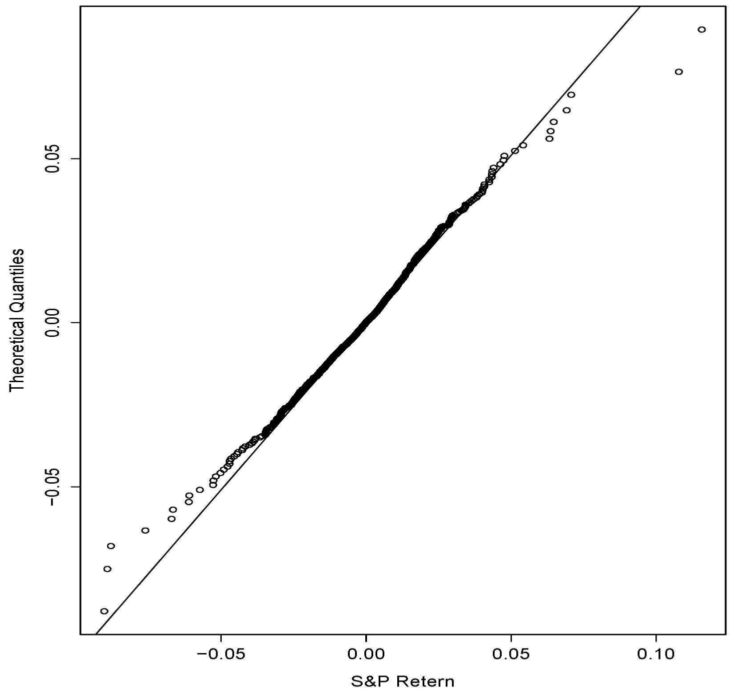


Figure 4 –: GDL probability plot.
GDL probability plot of SP 500 Returns. The reference line passes through the first and third quartiles.

## 5. Conclusion

A generalization of the double Lomax distribution has been proposed and some of its properties have been obtained, the symmetry and flexible heavy-tailedness of this distribution make it reasonable for modeling data in which small changes occur less frequently around origin and more likely within the heavy tails.

## References

A. Armagan, D. B. Dunson, J. Lee (2013). *Generalized double pareto shrinkage.* Statistica Sinica, 23, no. 1, p. 119.

P. Bindu (2011). *Estimation of p $(x > y)$ for the double lomax distribution.* In *Probstat forum.* vol. 4, pp. 1–11.

P. Bindu, K. Sangita (2015). *Double lomax distribution and its applications.* Statistica, 75, no. 3, p. 331.

M. Freimer, G. S. Mudholkar (1992). *An analogue of the chernoff-borovkov-utev inequality and related characterization.* Theory of Probability & Its Applications, 36, no. 3, pp. 589–592.

R. Korwar (1991). *On characterizations of distributions by mean absolute deviation and variance bounds.* Annals of the Institute of Statistical Mathematics, 43, no. 2, pp. 287–295.

K. Lomax (1954). *Business failures: Another example of the analysis of failure data.* Journal of the American Statistical Association, 49, no. 268, pp. 847–852.

D. Luethi, W. Breymann (2011). *ghyp: A package on the generalized hyperbolic distribution and its special cases.* R package version, 1, no. 5.

C. Necula, *et al.* (2009). *Modeling heavy-tailed stock index returns using the generalized hyperbolic distribution.* Romanian Journal of Economic Forecasting, 10, no. 2, pp. 118–131.

B. Pfaff (2012). *Financial risk modelling and portfolio optimization with R.* John Wiley & Sons, United Kingdom.

D. Ruppert (2011). *Statistics and data analysis for financial engineering.* Springer, New York.

C. Shannon (1948). *A mathematical theory of communication, bell syst. tech. j., 27: 376-423; 623-656. discrepancy and integration of continuous functions.* J. of Approx. Theory, 52, pp. 121–131.

R. D. C. Team, *et al.* (2011). *A language and environment for statistical computing.* R foundation for statistical computing, Vienna, Austria.

D. Wuertz, Y. Chalabi, M. Miklovic (2009). *fgarch: Rmetrics-autoregressive conditional heteroskedastic modelling.* R package version, 2110.

Yahoo (2007). *finance.* http://finance.yahoo.com. Accessed 22 November 2016.

Summary

A new probability distribution from the polynomial family has been proposed for modeling heavy-tailed data that are continuous on the whole real line. we have derived some general properties of this distribution and applied it on several data sets of U.S stock market daily returns. The introduced model is symmetric and leptokurtic, it outperforms the peer distributions used for the given data from perspective of information criteria suggesting a new potential candidate for modeling data exhibiting heavy tails.

*Keywords:* Heavy tailed distribution; Polynomial tails; Leptokurtic; Daily returns.