ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The Generalized Lambda Distribution as an Alternative Model to Financial Returns

Yohan Chalabi, David Scott and Diethelm Würtz

No. 2009-01

Rmetrics

ETH Econohysics Working and White Papers Series
Online at https://www.rmetrics.org/WhitePapers

The Generalized Lambda Distribution as an Alternative to Model Financial Returns

Yohan Chalabi,
Institut für Theoretische Physik,
HIT K21.2,
Wolfgang-Pauli-Strasse 27
8093 Zürich, Switzerland
chalabi@phys.ethz.ch

David J Scott,
Department of Statistics,
University of Auckland,
PB 92019,
Auckland,
New Zealand,
d.scott@auckland.ac.nz

and

Diethelm Würtz,
Institut für Theoretische Physik,
HIT K32.3,
Wolfgang-Pauli-Strasse 27
8093 Zürich, Switzerland
wuertz@phys.ethz.ch

Abstract:

We investigate the generalized lambda distribution with infinite support as an alternative distribution for modeling financial return series with power law tails. We derive expressions for the distribution, for random number generation, and for financial risk measures including value at risk, expected shortfall and tail indices. We introduce a new robust moment approach for the estimation of the distribution parameters based on the median, interquartile range, Bowley's skewness and Moors' kurtosis. In addition using a Monte Carlo approach we explore the use of several estimation approaches including maximum log likelihood, maximum product spacing, goodness of fit testing, and histogram binning. A new four-parameter parameterization allows an intuitive interpretation based on the asymmetry and the tail behaviour of the distribution. We also introduce a new method of obtaining parameter estimates in which the data is standardized to have zero median and unit interquartile range and then a generalized lambda distribution with zero median and unit interquartile range is fitted to the data. This reduces the number of parameters to two allowing for more efficient parameter estimation.

Keywords: Generalized lambda Distribution; Parameter Estimation, Maximum Likelihood, Maximum Product Spacings, Goodness-of-Fit, Histogram Binning, Financial Returns, Tail Index, Value at Risk, Shortfall Risk.

1 Introduction

The results in this paper are a contribution to knowledge in two areas. First of all we present new methods for fitting the generalized lambda distribution to data. Secondly, we show the utility of the generalized lambda distribution in the area of finance where up until now, very little investigation of it has been undertaken.

The four-parameter generalized lambda distribution (GLD) family is known for its high flexibility, producing distributions with a huge range of different shapes. The lambda distribution originated with Tukey [1962], and was generalized by Filliben [1969, 1975], Joiner and Rosenblatt [1971], and Ramberg and Schmeiser [1972, 1974].

The properties of the GLD were studied in detail by Ramberg et al. [1979]. The monograph of Karian and Dudewicz [2000] summarizes the status of research until 2000. Since then several papers have been published which present and discuss different parameter estimation approaches to fitting empirical data to this distribution. Parameter estimation of the GLD is notoriously difficult through the rapid change of the distributional shapes by varying the parameters in the different regions of the two shape parameters. In particular, the support of the distribution can change with the value of the parameters from being the whole real line to an interval which is infinite it only one direction. It is our view that fitting should be carried out only in regions where such dramatic changes do not take place. Our experience is that otherwise fitting routines typically become stuck in local, not necessarily global optima.

Little attention appears to have been paid to the GLD family in the area of finance, with the exception of Corrado [2001] and Tarsitano [2004]. The methods we develop in this paper allow the GLD to be easily fitted to empirical data such as financial returns. Calculations of important financial risk measures such as the value at risk, expected shortfall and tail indices are shown to be very simple for the GLD. We are also able to realistically simulate data arising from returns from equities. We show that using the GLD we can simulate data which closely resembles returns of the equities comprising the NASDAQ-100.

2 The Generalized Lambda Distribution

Ramberg and Schmeiser [1974] introduced the four parameter GLD defined by the quantile function

$$F^{-1}(p|\lambda) = F^{-1}(p|\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2} \tag{1}$$

where p are the probabilities, $p \in [0, 1]$, λ_1, λ_2 are the location and scale parameters, and λ_3, λ_4 are shape parameters jointly related to the strengths of the lower and upper tails, respectively. In the limiting case $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = \lambda_4 = \lambda$ we obtain Tukey's lambda distribution.

According to Ramberg and Schmeiser [1974] there exist six regions in which the shape parameters can lie in which the shapes of the GLDs are similar, see Figure 1.

The range of the parameters and the left and right support of the distributions are listed in Table 1. We see that the range can change abruptly when moving from one to another, for example from region 1 to region 5.

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

Table 1: The range of the GLD parameters and the minimum and maximum values corresponding to the labeling of the regions given in Figure 1.

The probability density function of the GLD at the point $x = F^{-1}(p)$ is given by

$$f(x) = f(F^{-1}(p)) = \frac{\lambda_2}{\lambda_3 p^{\lambda_3 - 1} + \lambda_4 (1-p)^{\lambda_4 - 1}} \tag{2}$$

Note that valid parameters for λ must fulfill the conditions

$$f(x) \geq 0 \tag{3}$$

$$\int f(x) dx = 1 \tag{4}$$

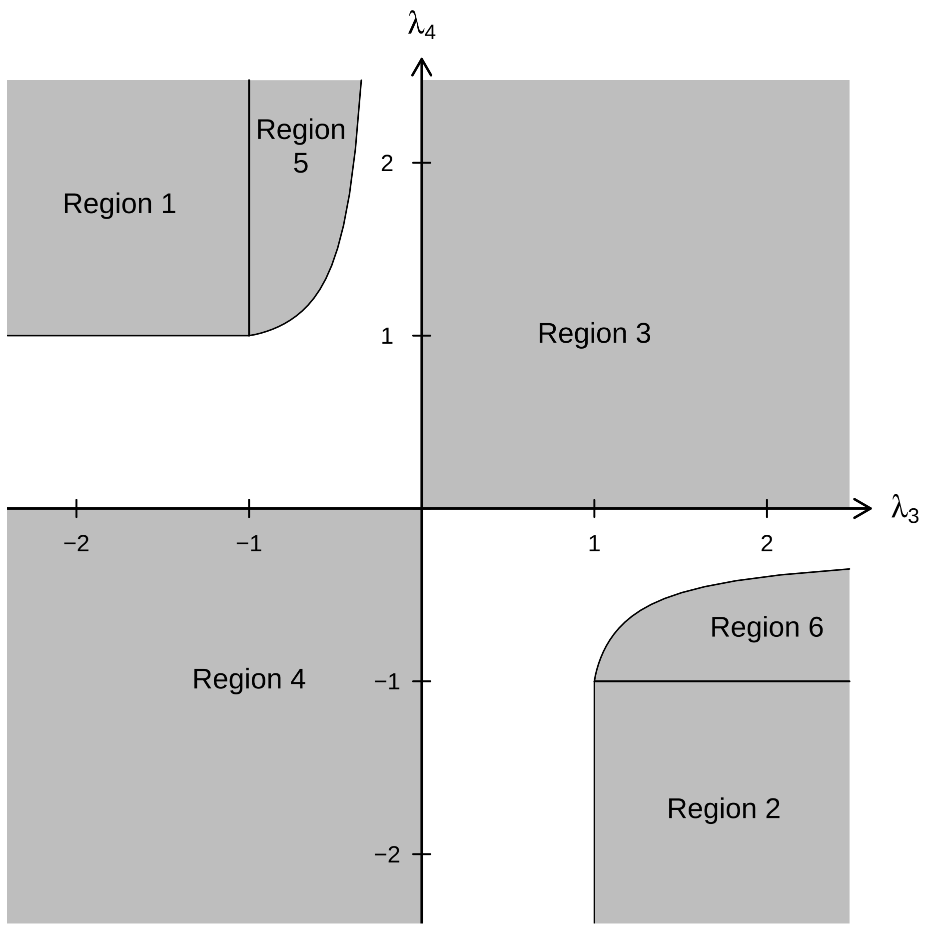

Figure 1: Graph of the parameter regions of the GLD in the λ_3 vs. λ_4 space that produce proper statistical distributions according to equations (3) and (4). The graph is as given in King and MacGillivray [1999] with the notation introduced by Ramberg et al. [1979].

These conditions determine the boundaries of the different parameter regions displayed in Figure 1. (See Ramberg and Schmeiser [1974] or Karian and Dudewicz [2000].)

3 GLDs for Modeling Financial Returns

Region 4 which defines the parameters in the third quadrant of the λ_3, λ_4 region makes the GLD interesting as an alternative skewed and heavily tailed candidate to the alpha-stable or Student-t distributions for modeling financial return series. In the following we will concentrate on the GLDs in this specific parameter region.

3.1 Stylized Facts

For negative shape parameters (and only in this parameter range) both the left and the right support of the distributions are infinite. The distributions are unimodal. When both shape parameters are equal the GLDs are symmetric about λ_1. The k-th moment only exists if $\min(\lambda_3, \lambda_4) > -1/k$ (Karian and Dudewicz [2000], Corollary 2.1.10).

Figure 2 shows the density and probability functions for the GLD with a fixed right tail parameter, λ_4 and a varying left tail parameter λ_3.

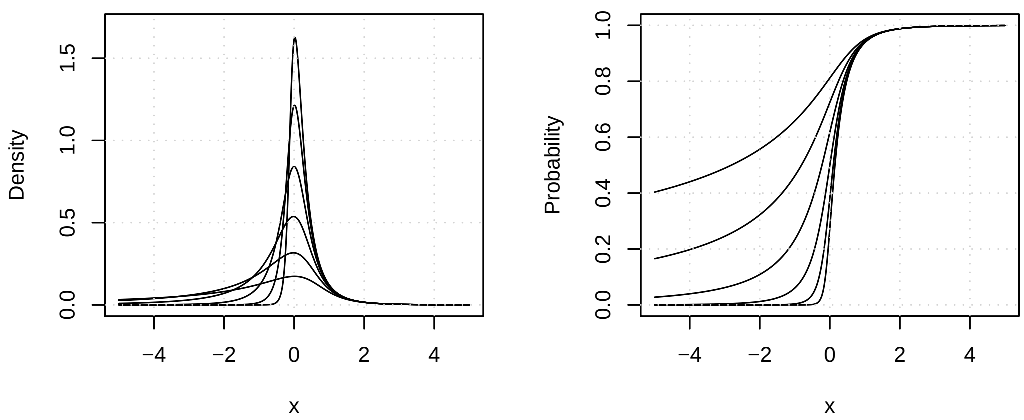

Figure 2: Density and probability function for the GLD in parameter region 4. The right tail is fixed at $\lambda_4 = -1/4$ and the left tail varies in powers of 2 in the range $\{-1/8, -1/4, -1/2, -1, -2\}$.

3.2 Tail Behaviour

From equation (1), we have

$$x = \lambda_1 + \frac{F(x)^{\lambda_3} - [1 - F(x)]^{\lambda_4}}{\lambda_2}$$

so that as $x \to \infty$

$$(1 - F(x))^{\lambda_4} \sim -\lambda_2 x$$

and hence

$$\frac{1 - F(tx)}{1 - F(x)} \sim t^{1/\lambda_4} \quad \text{as } x \to \infty \tag{5}$$

Thus the upper tail of the distribution function of the GLD is regularly varying at $+\infty$ with index $-1/\lambda_4$ (see **?**, p.129). A similar argument shows that the lower tail of the GLD is regularly varying at $-\infty$ with index $-1/\lambda_3$. Note that in the case we are considering both λ_3 and λ_4 are negative.

Also from the monotone density theorem, given as Theorem A3.7 in **?**, we have for the density function $f(x)$

$$f(x) \sim \frac{1}{\lambda_4}(-\lambda_2 x)^{1/\lambda_4 - 1} \quad \text{as } x \to \infty \tag{6}$$

with the analogous result holding as $x \to -\infty$.

The forgoing analysis shows the left and right tail probabilities are asymptotically power laws in region 4 of the GLD power space. If $\lambda_4 < -1$ the right tail density is asymptotically a power law also, as is the left tail if $\lambda_4 < -1$.

3.3 Value at Risk

Value at Risk, VaR, and expected shortfall risk, ES, are related to the quantiles of the distribution. Since the quantile function takes a simple algebraic form we can easily calculate these two risk measures:

$$\begin{aligned} \mathrm{VaR}_\alpha &= F^{-1}(\alpha|\lambda) && (7)\\ &= \lambda_1 + \frac{\alpha^{\lambda_3} - (1-\alpha)^{\lambda_4}}{\lambda_2} && \\ & && (8)\\ \mathrm{E}S_\alpha &= \int_{-\infty}^{\mathrm{VaR}} x f(x|\lambda)dx = \int_0^\alpha F^{-1}(p|\lambda)dp && (9)\\ &= \lambda_1\alpha + \frac{1}{\lambda_2(\lambda_3+1)}\alpha^{\lambda_3+1} + \frac{1}{\lambda_2(\lambda_4+1)}\left[(1-\alpha)^{\lambda_4+1} - 1\right] && \end{aligned}$$

4 Robust Moment Matching

The original method suggested by Ramberg et al. [1979] for fitting GLD distributions used the method of moments, matching the first four moments. However this approach can only sensibly be applied in regions where the first four moments exist. In region 4 this is the parameter range $\min(\lambda_3, \lambda_4) < -1/4$. Unfortunately this range excludes the major part of region 4. In addition parameter estimation based on sample moments, particularly third and fourth moments, is highly sensitive to outliers.

As an alternative Karian and Dudewicz [1999] considered a quantile approach for parameter estimation using four sample quantile statistics. In their approach all four statistics depend on all four λ parameters and one has to solve a system of four nonlinear equations to obtain the parameter estimates.

4.1 Robust Moments

As an alternative to these approaches we investigated matching of robust moments using the sample median μ_r, the interquartile range σ_r, the robust skewness ratio s_r of Bowley [1920], and the robust kurtosis ratio κ_r of Moors [1988]. These robust moments are defined by

$$\begin{aligned}
\mu_r &= \pi_{1/2} \\
\sigma_r &= \pi_{3/4} - \pi_{1/4} \\
s_r &= \frac{\pi_{3/4} + \pi_{1/4} - 2\pi_{2/4}}{\pi_{3/4} - \pi_{1/4}} \\
\kappa_r &= \frac{\pi_{7/8} - \pi_{5/8} + \pi_{3/8} - \pi_{1/8}}{\pi_{6/8} - \pi_{2/8}}
\end{aligned} \tag{10}$$

The subscript r in these definitions indicates they are robust versions of the moments. For a detailed discussion of Bowley's skewness and Moors' kurtosis statistics see Kim and White [2004].

The obvious estimators of the robust moments are then

$$\begin{aligned}
\hat{\mu}_r &= p_{1/2} \\
\hat{\sigma}_r &= p_{3/4} - p_{1/4} \\
\hat{s}_r &= \frac{p_{3/4} + p_{1/4} - 2p_{2/4}}{p_{3/4} - p_{1/4}} \\
\hat{\kappa}_r &= \frac{p_{7/8} - p_{5/8} + p_{3/8} - p_{1/8}}{p_{6/8} - p_{2/8}}
\end{aligned} \tag{11}$$

where p_q indicates the sample qth quantile and the hat that the statistic is a sample quantity.

Defining $S_{\lambda_3,\lambda_4}(p)$ as

$$S_{\lambda_3,\lambda_4}(p) = S(p|\lambda_3, \lambda_4) = p^{\lambda_3} - (1-p)^{\lambda_4} \tag{12}$$

we have

$$F^{-1}(p|\lambda) = \lambda_1 + \frac{S_{\lambda_3,\lambda_4}(p)}{\lambda_2}$$

and we see that the population robust skewness and kurtosis depend only on the parameters λ_3 and λ_4 and not on λ_1 or λ_2. Specifically

$$\begin{aligned}
\mu_r &= \lambda_1 + \frac{S_{\lambda_3,\lambda_4}(1/2)}{\lambda_2} \\
\sigma_r &= -\frac{S_{\lambda_3,\lambda_4}(3/4) - S_{\lambda_3,\lambda_4}(1/4)}{\lambda_2} \\
s_r &= \frac{S_{\lambda_3,\lambda_4}(3/4) + S_{\lambda_3,\lambda_4}(1/4) - 2S_{\lambda_3,\lambda_4}(1/2)}{S_{\lambda_3,\lambda_4}(3/4) - S_{\lambda_3,\lambda_4}(1/4)} \\
\kappa_r &= \frac{S_{\lambda_3,\lambda_4}(7/8) - S_{\lambda_3,\lambda_4}(5/8) + S_{\lambda_3,\lambda_4}(3/8) - S_{\lambda_3,\lambda_4}(1/8)}{S_{\lambda_3,\lambda_4}(6/8) - S_{\lambda_3,\lambda_4}(2/8)}
\end{aligned} \tag{13}$$

Given these definitions we first use the observed values $\hat{s}_r$ and $\hat{\kappa}_r$ of the robust skewness and kurtosis to obtain estimates $\hat{\lambda}_3$ and $\hat{\lambda}_4$ of λ_3 and λ_4 by inverting the nonlinear system of equations given by

$$\begin{aligned}
\hat{s}_r &= s_r(\hat{\lambda}_3, \hat{\lambda}_4) \\
\hat{\kappa}_r &= \kappa_r(\hat{\lambda}_3, \hat{\lambda}_4).
\end{aligned} \tag{14}$$

The remaining two parameters λ_1, λ_2 are then estimated as

$$\hat{\lambda}_2 = -\frac{S_{\hat{\lambda}_3,\hat{\lambda}_4}(1/2)}{\hat{\sigma}_r} \tag{15}$$

and then

$$\hat{\lambda}_1 = \hat{\mu}_r - \frac{S_{\hat{\lambda}_3,\hat{\lambda}_4}(3/4) - S_{\hat{\lambda}_3,\hat{\lambda}_4}(1/4)}{\hat{\lambda}_2} \tag{16}$$

The beauty of this approach is that we decouple the shape parameters from the location and scale parameters, and we only need to solve a non-linear system of two equations rather than the system of four equations required by previous methods. Beside being simpler, this also allows the use of a lookup method for obtaining initial estimates of λ_3 and λ_4 given the sample robust skewness and kurtosis. We thus have the possibility of very quickly obtaining starting values for any numerical approach to determining maximum likelihood or other estimators of the parameters of the GLD.

4.2 Parameter Estimation from Robust Moments

We used the method developed in the previous section for the equities from the NASDAQ-100 index. The NASDAQ-100 Index includes 100 of the largest US domestic and international non-financial securities listed on the Nasdaq Stock Market based on market capitalization, see Table 2. The index reflects the share prices of companies across major industry groups including computer hardware and software, telecommunications, retail/wholesale trade and biotechnology. Although it does not contain securities of financial companies we expect that the listed equities are heavy tailed, and thus are good candidates for modeling their distributions with the GLD.

Since the index composition has changed over time, the lengths of the time series vary. We have not included the influence of this effect in our investigation. To get an idea on the length of the time series in the NASDAQ-100 data set, see Figure 3.

AAPL	ADBE	ADP0	ADSK	AKAM	ALTR	AMAT	AMGN	AMZN	APOL
ATVI	BBBY	BIDU	BIIB	BRCM	CA00	CELG	CEPH	CHKP	CHRW
CMCSA	COST	CSCO	CTAS	CTSH	CTXS	DELL	DISH	DTV0	EBAY
ERTS	ESRX	EXPD	EXPE	FAST	FISV	FLEX	FLIR	FSLR	FWLT
GENZ	GILD	GOOG	GRMN	HANS	HOLX	HSIC	IACI	ILMN	INFY
INTC	INTU	ISRG	JAVA	JBHT	JNPR	JOYG	KLAC	LBTYA	LIFE
LINTA	LLTC	LOGI	LRCX	MCHP	MICC	MRVL	MSFT	MXIM	NIHD
NTAP	NVDA	NWSA	ORCL	ORLY	PAYX	PCAR	PDCO	PPDI	QCOM
RIMM	ROST	RYAAY	SBUX	SHLD	SIAL	SPLS	SRCL	STLD	STX0
SYMC	TEVA	URBN	VRSN	VRTX	WCRX	WYNN	XLNX	XRAY	YHOO

Table 2: NASDAQ Symbols. The index series were downloaded from `finance.yahoo.com` as of November, 2009.

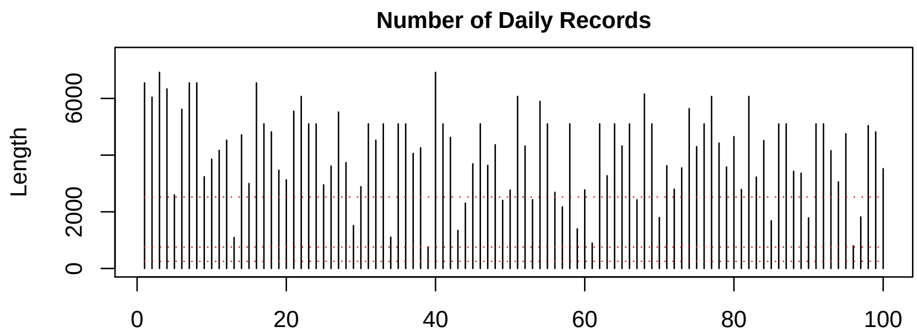

Figure 3: NASDAQ-100 record lengths for each index equity. The index series are in alphabetical order as listed in Table 2. The dotted horizontal lines mark the 1, 3, and 10 year record lengths.

We downloaded the financial index series for the NASDAQ-100 from Yahoo's finance web portal. We selected the adjusted close as index data. Missing data records due to holidays have been linearly interpolated in the index.

First we estimated the empirical moments $\hat{s}_r$, and $\hat{\kappa}_r$ from which we then obtained estimates for the parameters λ_3 and λ_4. The relationships between the two shape parameters and the empirical robust skewness and kurtosis are shown in Figure 4, and the caption explains the content of the graphs. Note that all equities for which either λ_3 or λ_4 is less than -0.25 exhibit finite sample variance and sample kurtosis, between -0.25 and -0.50 the variance exists and the kurtosis diverges, and above -0.50, both variance and kurtosis are infinite. The mean of the estimated values for the robust skewness is 0.0319, for the robust kurtosis 1.56. The associated value for λ_3 is -0.364, and for λ_4 -0406. These points are marked on the graphs as a light coloured point in the midst of the points for individual equities.

5 Parameter Fitting

To fit the parameters we considered four different approaches, fitting a histogram, minimizing a goodness of fit statistic, maximum log-likelihood, and the maximum product spacing. The last method (see Cheng and Amin [1983] or Ranneby [1984]) has not previously been used for the estimation of GLD parameters.

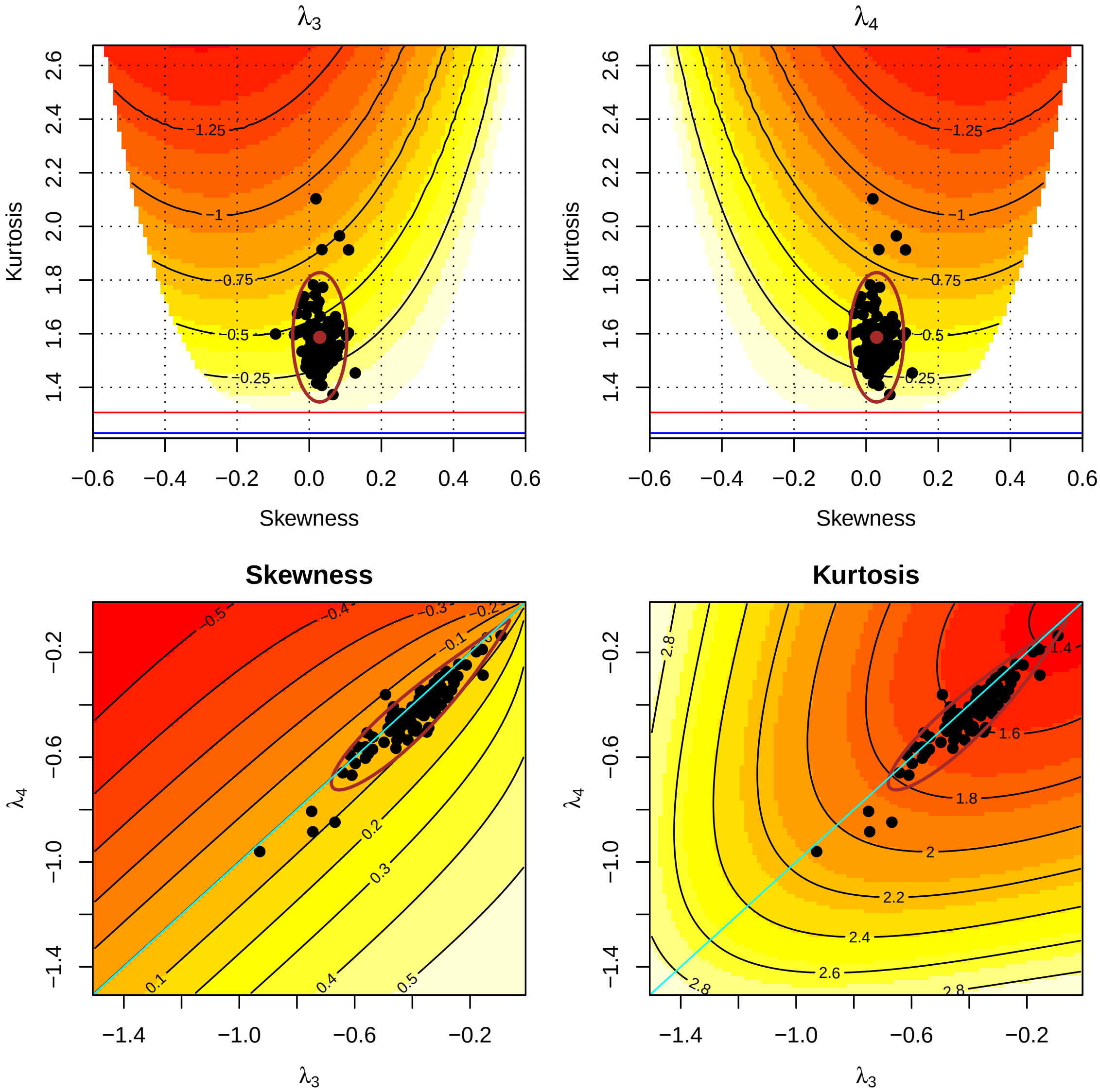

Figure 4: Estimates of λ_3 and λ_4 for the NASDAQ-100 equities derived from the robust quantile estimates. The upper two graphs show scatterplots for the robust sample skewness and kurtosis on top of an image and contour plot for the parameter estimates $\hat{\lambda}_3$ and $\hat{\lambda}_4$. The correlation ellipse contains 90% of the data. The lower two graphs show the inverted map plotting the parameter estimates $\hat{\lambda}_3$ versus $\hat{\lambda}_4$. Here the contours show constant levels of the skewness and kurtosis. The dots represent the NASDAQ equities and the closed line the correlation transformed ellipse. The diagonal line represents the case of symmetric GLDs.

5.1 Histogram Approaches

The histogram approach is very appealing and simple. The empirical data are binned in a histogram and the resulting probabilities at the midpoints of the histogram bins are fitted to the true GLD density. This approach was considered by Su [2005] for the GLD. The central question is then how many bins should we use to obtain the best fit? We investigated three different methods to estimate the number of bins. Let n denote the sample size and b_n the number of bins for that sample size.

- *Sturges Breaks:* Due to Sturges [1926], this is the default method used in many software programs to calculate the number of histogram bins. Sturges' formula computes bin sizes from the range of the data and can perform quite poorly if the sample size is small, e.g. n less than 30. For this method
$$b_n = \left\lceil \log_2(n+1) \right\rceil . \tag{17}$$
The brackets represent the ceiling function, which indicates that b_n will be rounded up to the next integer value.

- *Scott Breaks:* Scott's choice (see Scott [1979]) is derived from the normal distribution and relies on the estimate of the standard error. This is a more flexible approach compared to the fixed number of bins used by Sturges. The number of bins is calculated from the formula
$$\begin{aligned} h &= 3.49 \hat{\sigma} n^{1/3} \\ b_n &= \left\lceil \frac{\max(x) - \min(x)}{h} \right\rceil . \end{aligned} \tag{18}$$
When all observations are equal, $h = 0$, and we take $b_n = 1$.

- *Freedman and Diaconis Breaks:* The Freedman-Diaconis choice (Freedman and Diaconis [1981]) is a robust selection method based on the range and the inter-quartile range of the data. The optimal number of bins can be computed as
$$\begin{aligned} h &= \frac{\pi(0.75) - \pi(0.25)}{n^{1/3}} \\ b_n &= \left\lceil \frac{\max(x) - \min(x)}{h} \right\rceil . \end{aligned} \tag{19}$$
If the IQR is zero, then the median absolute deviation is used for h.

The Sturges breaks do not use any information concerning the form of the underlying distribution and the histogram bins are often far from optimally chosen. The Scott approach relies on the distribution being shaped like a normal distribution and thus for the GLD we can expect histogram bins which are not optimally chosen with respect to the tails. The Freedman-Diaconis approach seems to be the most promising one to create histograms from random variates drawn from the heavy tailed GLD, since it is a robust method based on the inter-quartile range.

5.2 GoF Approaches

To test the goodness of fit of a continuous cumulative distribution function we can use statistics based on the empirical distribution function, such as the Kolmogorov-Smirnov, Cramér-von Mises, or Anderson-Darling statistics. These statistics measure the discrepancy between the hypothetical GLD distribution and the empirical distribution.

To use these statistics as parameter estimators we minimize them with respect to the unknown parameters of the distribution. For example, it is known from the generalized Pareto distribution that these approaches are able to deal successfully with the estimation of the parameters even when the maximum likelihood method fails, as shown by Luceño [2006], who calls this approach the method of maximum goodness of fit. Definitions of the the three goodness of fit statistics which we use, and computational forms for them, are also given in the paper of Luceño [2006]. Consider a random sample of n *iid* observations $\{x_1, x_2, \ldots, x_n\}$ and let $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ denote the order statistics of the sample. Write $F_i = F(x_{(i)})$.

- *Kolmogorov-Smirnov Statistic:* The KS statistic measures the maximum distance
$$D_n = \frac{1}{2n} + \max_{1 \le i \le n} \left| F_i - \frac{1 - i/2}{n} \right| \tag{20}$$
which leads to a non-continuous objective function for the optimization.

- *Cramér-von Mises Statistic:* The CvM statistic uses the mean squared differences leading to a continuous objective for the parameter optimization
$$W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(F_i - \frac{1 - i/2}{n} \right)^2 \tag{21}$$

- *Anderson-Darling Statistic:* The AD statistic is a tail weighted statistic which gives more weights to the tails and less to the central part of the distribution which renders the AD more suitable for use with the heavy tailed GLD

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1)(\ln(F_i) + \ln(1 - F_{n+1-i}) \tag{22}$$

5.3 MLE and MPS Approaches

The Kullback-Leibler (KL) divergence (Kullback and Leibler [1951] and Kullback [1959]), which is also known under the names information divergence, information gain, or relative entropy, measures the difference between two probability distributions P and Q. Typically P represents the empirical distribution, and Q comes from a theoretic model. The KL divergence allows for the interpretation of many other measures in information theory. Two of them are the maximum log likelihood estimator and the maximum product spacings estimator which we will use to fit the parameters of GLD.

- *Maximum Log-Likelihood Estimation:*

  The maximum log-likelihood method was introduced by Fisher [1922]. For a more recent review we refer to Aldrich [1997]. Consider a random sample of n *iid* observations $\{x_1, x_2, \ldots, x_n\}$ drawn from the GLD with parameters λ. Then the maximum value of the log-likelihood function $\mathcal{L}$ defined by

$$\mathcal{L}(\lambda) = \sum_{i=1}^{n} \ln\left(f(x_i|\lambda)\right) \tag{23}$$

  returns optimal estimates for the parameters λ of the GLD. The maximum of this expression can be found numerically using a non-linear optimizer allowing for linear constraints.

- *Maximum Product Spacings:*

  The maximum likelihood method may break down in various cases, for example where the support depends on the parameters to be estimated, see for example Cheng and Amin [1983]. Under these circumstances the method of maximum product spacing, MPS, introduced separately by Cheng and Amin [1983] and Ranneby [1984] may be more successful. Consider the order statistics $\{x_{(1)}, x_{(2)}, \ldots, x_{(n)}\}$ of the observations $\{x_1, x_2, \ldots, x_n\}$ and compute their spacings or gaps at adjacent points

$$D(x_{(i)}|\lambda) = F\left(x_{(i)}|\lambda\right) - F\left(x_{(i-1)}|\lambda\right) \tag{24}$$

then the maximum of the max log spacing function $\mathcal{D}$

$$\mathcal{D}(\lambda) = \sum_{i=2}^{n} \ln\left(D(x_{(i)}|\lambda)\right) \tag{25}$$

returns optimal estimates for the parameters of the GLD. The maximum of this expression can be found numerically as in the case of the MLE using a non-linear optimizer allowing for linear constraints.

5.4 Parameter Estimation

We implemented the histogram, GOF, MLE and MPS parameter estimation approaches in the statistical software environment R. We used the Friedman-Diaconis breaks for the histogram approach and the Anderson-Darling statistic for the goodness of fit approach. A graph showing the results of these four parameter estimation methods is shown in Figure 5 for the shares of Google.

We observe that there is very little difference discernable between the goodness of fit (AD) and MPS approaches. The MLE fit differs from AD and MPS in the tails by a small amount. The tail fit for the histogram (FD) approach is substantially different in the upper tail.

6 Performance Comparison

Which estimator is best? This is not clear a priori, since in finance distributions are typically heavy-tailed, sometimes with small sample sizes, and often the financial return series have outlying data points.

To investigate the performance of the individual estimators depending on the heaviness of the tails, the sample size and outlying jumps, we generate a large set of randomly variables simulating financial return series as observed typically on stock markets.

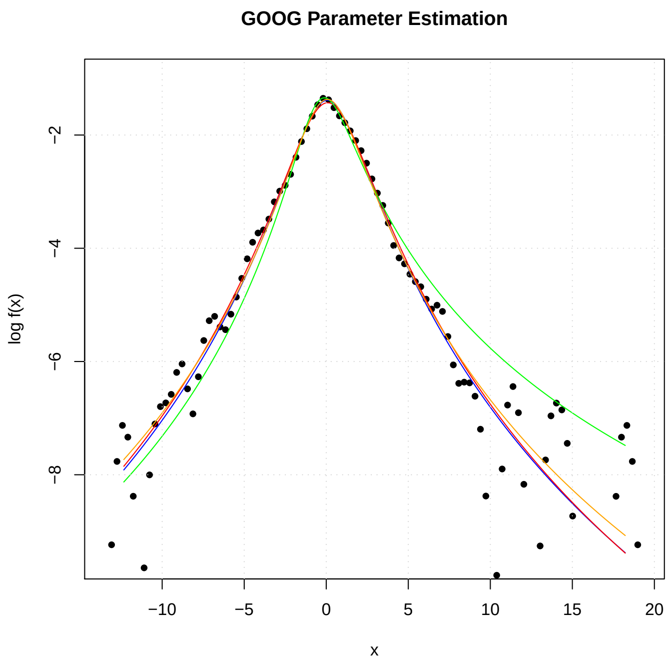

Figure 5: Parameter estimation for the Google equity. The graph shows the results from the MLE (blue), MPS (red), AD (orange), and FD (green) approaches. The full lines are drawn from the fitted distribution function and the points are taken from a kernel density estimate of the simulated series.

6.1 MC Simulation of Parameters

Simulating a set of λ values we have to keep in mind that the four λ parameters may be highly correlated. The correlation plot for the 100 NASDAQ equities shows that the inter-quartile range and the tail related shape parameters are highly correlated. Introducing a new set of parameters $\{\lambda_1, \lambda_2, \delta, \beta\}$ where $\delta = \lambda_3 - \lambda_4$, and $\beta = \lambda_3 + \lambda_4$, we observe from Figure 6 that the only substantial correlations are λ_1 with δ and λ_2 with β.

To generate artificial time series parameters by a Monte Carlo simulation we proceed as follows:

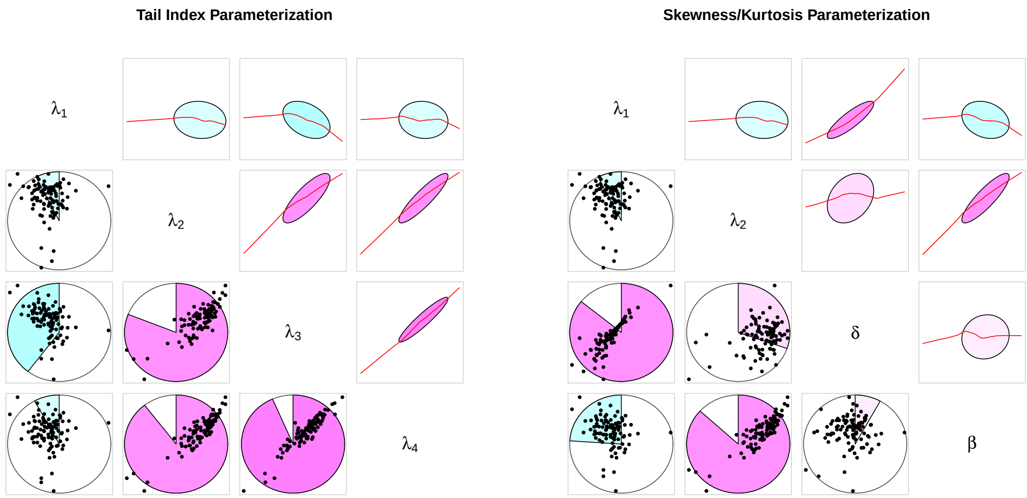

Figure 6: Correlation of the distributional parameters. On the left hand side for the $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ parameterization, and on the right hand side for the alternative $\{\lambda_1, \lambda_2, \delta, \beta\}$ parameterization.

- *First*, we estimate the median, the inter-quartile range, and the robust skewness and kurtosis parameters from the 100 NASDAQ equities, invert equations (14) and use equations (15) and (15) to get the sample λ_1, λ_2, δ and βs.

- *Second*, we compute from the parameters λ_1, λ_2, β, and δ density estimates using the smoothing spline ANOVA approach of **?** and **?**, see Figure 7.

- *Third*, we estimate the dependency structures of λ_1 vs. δ and λ_2 vs β from two bivariate Gaussian copulas, see Figure 8.

- *Fourth*, finally we generate random variates for the probabilities from the copulas and compute from the marginal distributions the parameters λ_1, λ_2, δ, and β. λ_3 and λ_4 are recalculated from δ and β.

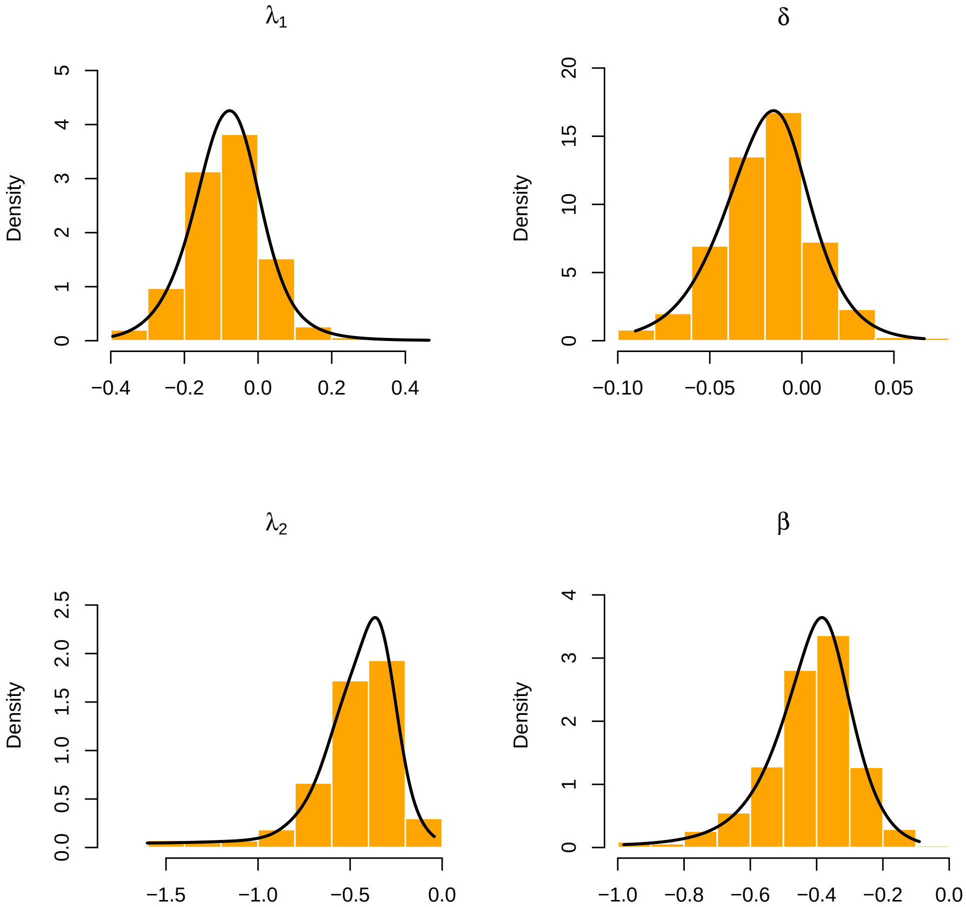

Figure 7: The marginal distribution functions an the ANOVA spline fits for the marginal distributions of parameters.

The outcome of this Monte Carlo approach is shown in Figures 7 and 8. Figure 7 shows the histograms of the parameters $\{\lambda_1, \lambda_2, \delta, \beta\}$ as obtained from the 100 NASDAQ equities together with the fitted distributions from which we draw the random variates. That the parameters λ_2 and β preserve the observed dependence structure is shown in Figure 8 where we have plotted the sample correlation together with the Gaussian copula.

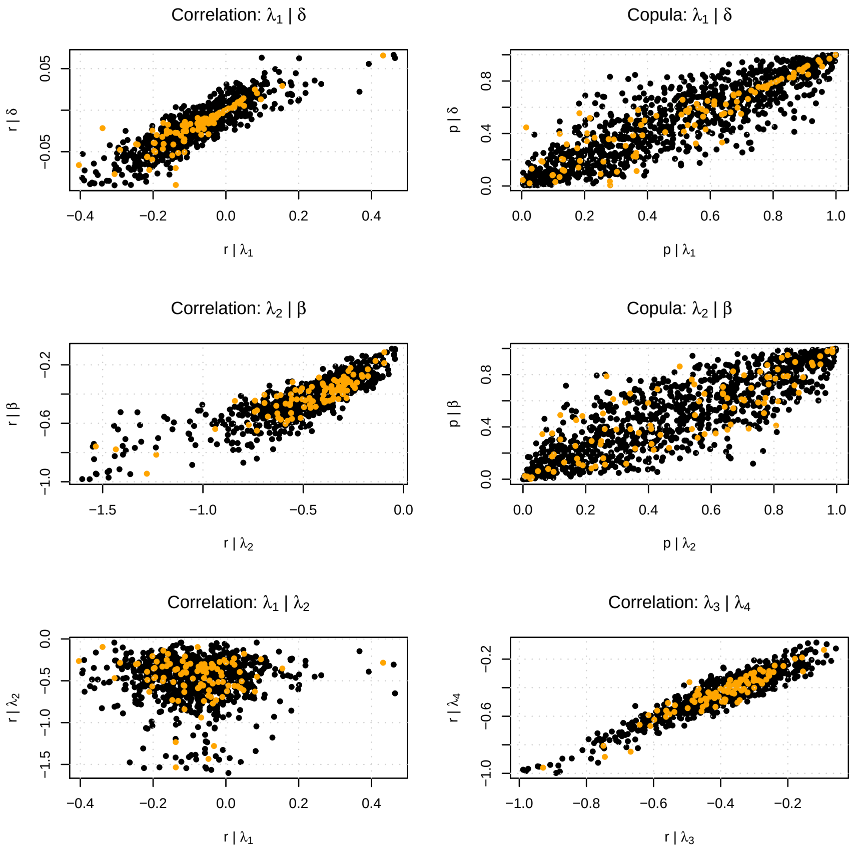

Figure 8: Simulation of distributional parameters with the same statistical properties as those of the NASDAQ-100 equity returns. The upper row shows the correlation and copula dependence structure for the λ_1, δ parameter pair, and the row in the middle the data for the λ_2, β parameter pair. In the lower row the correlation is shown for λ_1, λ_2 and λ_3, λ_4 pairs derived from the simulated δ and β parameters.

All Simulations

Large Samples N=1000

lambda 1

Distance	mle	mps	ad	fd	rob
euclid	0.404	1.000	0.420	0.577	0.509
bc	0.617	0.728	0.636	0.862	0.656
cor	0.259	0.585	0.274	0.397	0.338
tau	0.481	0.488	0.483	0.574	0.483
mut	0.284	0.606	0.297	0.374	0.395

lambda 2

Distance	mle	mps	ad	fd	rob
euclid	0.063	0.993	0.071	0.141	0.337
bc	0.062	0.107	0.071	0.137	0.078
cor	0.045	0.832	0.055	0.180	0.505
tau	0.216	0.230	0.240	0.396	0.228
mut	0.102	0.923	0.107	0.221	0.570

lambda 3

Distance	mle	mps	ad	fd	rob
euclid	0.422	0.572	0.479	1.000	0.448
bc	0.047	0.052	0.052	0.108	0.048
cor	0.048	0.077	0.060	0.199	0.053
tau	0.212	0.217	0.235	0.408	0.217
mut	0.077	0.088	0.091	0.217	0.081

lambda 4

Distance	mle	mps	ad	fd	rob
euclid	0.422	0.572	0.479	1.000	0.448
bc	0.045	0.060	0.051	0.104	0.050
cor	0.053	0.663	0.067	0.209	0.254
tau	0.222	0.234	0.248	0.428	0.233
mut	0.090	0.829	0.104	0.227	0.217

Small Samples L=100

lambda 1

Distance	mle	mps	ad	fd	rob
euclid	0.422	0.572	0.479	1.000	0.448
bc	0.492	0.491	0.501	0.675	0.854
cor	0.177	0.177	0.173	0.238	0.311
tau	0.410	0.412	0.411	0.474	0.510
mut	0.190	0.190	0.202	0.201	0.267

lambda 2

Distance	mle	mps	ad	fd	rob
euclid	0.422	0.572	0.479	1.000	0.448
bc	0.049	0.053	0.056	0.122	0.116
cor	0.062	0.061	0.072	0.262	0.232
tau	0.243	0.242	0.279	0.482	0.468
mut	0.083	0.079	0.085	0.205	0.181

lambda 3

Distance	mle	mps	ad	fd	rob
euclid	0.422	0.572	0.479	1.000	0.448
bc	0.032	0.034	0.036	0.084	0.084
cor	0.104	0.104	0.132	0.378	0.346
tau	0.310	0.310	0.340	0.547	0.548
mut	0.104	0.106	0.126	0.235	0.262

lambda 4

Distance	mle	mps	ad	fd	rob
euclid	0.422	0.572	0.479	1.000	0.448
bc	0.032	0.033	0.036	0.082	0.077
cor	0.125	0.126	0.148	0.408	0.362
tau	0.348	0.349	0.373	0.582	0.535
mut	0.116	0.116	0.135	0.239	0.265

Heavy tailed

Large Samples N=1000

lambda 1

Distance	mle	mps	ad	fd	rob
euclid	0.599	0.595	0.595	0.754	1.000
bc	0.492	0.491	0.501	0.675	0.854
cor	0.177	0.177	0.173	0.238	0.311
tau	0.410	0.412	0.411	0.474	0.510
mut	0.190	0.190	0.202	0.201	0.267

lambda 2

Distance	mle	mps	ad	fd	rob
euclid	0.417	0.469	0.462	1.000	0.937
bc	0.049	0.053	0.056	0.122	0.116
cor	0.062	0.061	0.072	0.262	0.232
tau	0.243	0.242	0.279	0.482	0.468
mut	0.083	0.079	0.085	0.205	0.181

lambda 3

Distance	mle	mps	ad	fd	rob
euclid	0.379	0.399	0.435	1.000	0.968
bc	0.032	0.034	0.036	0.084	0.084
cor	0.104	0.104	0.132	0.378	0.346
tau	0.310	0.310	0.340	0.547	0.548
mut	0.104	0.106	0.126	0.235	0.262

lambda 4

Distance	mle	mps	ad	fd	rob
euclid	0.408	0.421	0.453	1.000	0.951
bc	0.032	0.033	0.036	0.082	0.077
cor	0.125	0.126	0.148	0.408	0.362
tau	0.348	0.349	0.373	0.582	0.535
mut	0.116	0.116	0.135	0.239	0.265

Small Samples L=100

lambda 1

Distance	mle	mps	ad	fd	rob
euclid	0.388	0.373	0.390	0.581	1.000
bc	1.643	1.646	1.625	1.991	2.601
cor	0.580	0.581	0.576	0.688	0.931
tau	0.713	0.718	0.717	0.796	0.835
mut	0.370	0.347	0.365	0.466	0.499

lambda 2

Distance	mle	mps	ad	fd	rob
euclid	0.415	0.571	0.455	0.920	1.000
bc	0.191	0.213	0.193	0.344	0.357
cor	0.419	0.411	0.409	0.568	0.576
tau	0.629	0.618	0.608	0.734	0.740
mut	0.319	0.303	0.319	0.305	0.436

lambda 3

Distance	mle	mps	ad	fd	rob
euclid	0.496	0.543	0.503	1.000	0.974
bc	0.119	0.121	0.120	0.234	0.250
cor	0.475	0.471	0.480	0.615	0.678
tau	0.648	0.649	0.679	0.787	0.839
mut	0.246	0.236	0.231	0.338	0.342

lambda 4

Distance	mle	mps	ad	fd	rob
euclid	0.529	0.586	0.505	0.966	1.000
bc	0.115	0.118	0.109	0.220	0.237
cor	0.547	0.551	0.569	0.748	0.710
tau	0.722	0.724	0.755	0.893	0.867
mut	0.329	0.297	0.282	0.288	0.343

Figure 9: Distance measures obtained from simulation of 1000 samples. The upper blocks consider large samples each of size 1000, lower blocks consider smaller samples each of 100. The left blocks show the statistics for all samples, the right blocks the results for heavy tailed samples.

6.2 Parameter Estimation

The 1000 samples generated for the parameters allow us to simulate random financial return samples x of any length L which correspond to daily recorded series of a period of approximately 4 years. For each simulated random series we have estimated the distributional parameters using the following estimators

- Maximum Log-Likelihood Estimator
- Maximum Product Spacings Estimator
- Anderson Darling Statistics Estimator
- Freedman-Diaconis Histogram Estimator

Since we know the true parameters we can compute performance measure statistics to compare their quality. To see how similar the Monte Carlo results for the different estimators are and to make a ranking for the estimators we have compared them based on several distance measures, as outlined below.

- Euclidean Distance: Calculates the pairwise Euclidean distances for all columns x_i, x_j of the input matrix. $d(x, y) = \sqrt{x_i^2 - y_i^2}$.
- Bray Curtis Distance: Calculates the normalized pairwise distances on a grid for all columns of the input matrix. $d(x, y) = |x_i - y_i|/(x_i + y_j)$.
- Correlation Distance: Calculates for all rows of the input matrix the pairwise Pearson correlation distance $d = 1 - \mathrm{cor}(x, y)$.
- Kendalls τ Distance: Calculates for all rows of the input matrix the pairwise Kendall τ correlation distance $d = 1 - \tau$.
- Mutual Information Distance: Each row of λ was divided into n binned groups and then the mutual information was computed, treating the data as if they were discrete. Then we used the transformation proposed by **?**, $d(x, y) = 1 - \sqrt{1 - \exp[-2I(x, y)]}$ where $I(x, y)$ is the mutual information.

We generated 1000 larger and smaller samples, the first of size 1000 and the second of size 100. We investigated all 1000 samples together, and in a second step we separated those with heavier tails, i.e. those where the estimated parameters indicated the data came from a distribution with infinite second and fourth moment.

7 Alternative Parameterizations

The original parameterization in the four lambda parameters characterizes GLDs with respect to the tails: λ_3 determines the shape of the lower tail, and λ_4 the shape of the lower tail. A descriptive name for this parameterization is the *Tail Index Parameterization* of the GLD.

7.1 Skewness-Kurtosis Parameterization

The alternative parameterization in parameters β and δ reflects the robust skewness and robust kurtosis of the distribution. A descriptive name for this parameterization is the *Skewness Kurtosis Parameterization* of the GLD. This parameterization allows us to draw a shape triangle similar to the shape triangle for the generalized hyperbolic distribution. (See for example ?.)

The shape triangle is drawn in Figure 10 together with the parameter estimates for the NASDAQ-100 listed equities. The shape triangle allows for a simple interpretation. The x-axis measures the asymmetry of the up and down movements of the equities, and the y-axis expresses the heaviness of the tails.

7.2 Standardized Parameterization

The representation of the parameters through robust moments allows us to standardize the distribution. This leads to a parameterization where the distribution has zero median and unit inter-quartile range, and two shape parameters. The standardized distribution takes the form

$$
\begin{aligned}
f(x|\lambda_3, \lambda_4) &= f(x|\lambda_1^\star, \lambda_2^\star, \lambda_3, \lambda_4) \\
\lambda_2^\star &= S_{\lambda_3,\lambda_4}(3/4) - S_{\lambda_3,\lambda_4}(1/4) \\
\lambda_1^\star &= -S_{\lambda_3,\lambda_4}(1/2)/\lambda_2
\end{aligned} \tag{26}
$$

where $S_{\lambda_3,\lambda_4}(p)$ is defined by equation (12) It is easily verified from eqn. () equation (13), that the median of $f(x)$ is zero and the inter-quartile range is equal one.

Note, the standardized parameterization can also be written using the parameters β and δ.

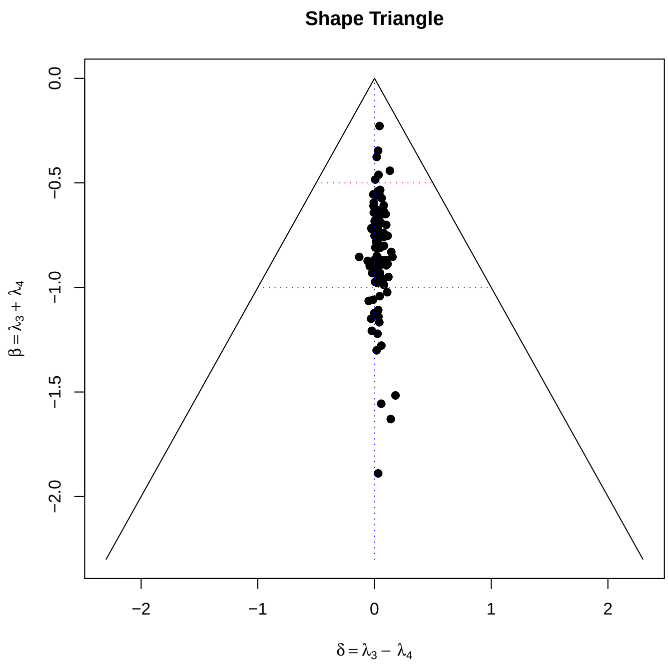

Figure 10: The graph shows the shape triangle. the parameters $\lambda_3 - \lambda_4$ plotted versus $\lambda_3 + \lambda_4$. The points in the shape triangle represent estimates for the the NASDAQ-100 listed equities. Shares on the right hand side of the blue line have an upward trend, those to left a downward trend. In the peak of the triangle above -0.5 the equities have finite variance and kurtosis, in the middle range between -1 and -0.5 the equities have infinite kurtosis but still finite variance. Below -1 both moments are infinite.

8 Conclusions

The generalized lambda distribution with infinite lower and upper support is an interesting alternative with power law tails to the stable and the Student-t distribution when modeling financial returns. The quantile function is easily expressed as a function of the probability and the four distributional parameters. From the quantile function the distribution and density function can be easily derived by simple inversions. Value at risk, expected shortfall, and the tail index can be expressed in simple formulas depending on the parameters. Introducing robust moments and defining the robust skewness and kurtosis we are able to separate the median and inter-quartile range. Then we are left with two nonlinear equations to be inverted to get an estimate parameters from the moments. The robust moment estimates can be used as a starting solution for several parameter estimators including maximum log likelihood, maximum product spacing, goodness of fit, and histogram binning approaches. An extensive Monte Carlo simulation shows, that the maximum log likelihood approach is the most reliable one also for small samples and heavy tails.

Acknowledgement

YC acknowledges financial support through a PhD grant given by Finance Online GmbH in Zurich. Part of this work was done in the PhD project of YC. DS is grateful for the hospitality of the Institute of Theoretical Physics, ETH Zürich, where much of this work was undertaken.

References

John Aldrich. R. A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3):162–176, August 1997.

A.L. Bowley. *Elements of Statistics.* Scribner, New York, 1920.

R. C. H. Cheng and N. A. K. Amin. Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3):394–403, 1983.

Charles J. Corrado. Option pricing based on the generalized lambda distribution. *Journal of Futures Markets*, 21(3):213–236, 2001.

J. J. Filliben. *Simple and robust estimation of the location parameter of a symmetric distribution.* PhD thesis, Princeton University, Princeton, New Jersey, 1969.

James J. Filliben. The probability plot correlation coefficient test for normality. *Technometrics*, 17(1):111–117, February 1975.

R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.

David Freedman and Persi Diaconis. On the histogram as a density estimator: L_2 theory. *Probability Theory and Related Fields*, 57(4):453–476, December 1981.

Brian L. Joiner and Joan R. Rosenblatt. Some properties of the range in samples from tukey's symmetric lambda distributions. *Journal of the American Statistical Association*, 66(334):394–399, June 1971.

Z.A. Karian and E.J. Dudewicz. *Fitting statistical distributions to data: Generalized lambda distribution and generalized bootstrap methods.* CRC Press, New York, 2000.

Zaven A. Karian and Edward J. Dudewicz. Fitting the generalized lambda distribution to data: a method based on percentiles. *Communications in Statistics - Simulation and Computation*, 28(3):793–819, 1999.

Juha Karvanen and Arto Nuutinen. Characterizing the generalized lambda distribution by l-moments. *Computational Statistics & Data Analysis*, 52 (4):1971–1983, January 2008.

Tae-Hwan Kim and Halbert White. On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1):56–73, March 2004.

Robert A. R. King and H. L. MacGillivray. A starship estimation method for the generalized λ distributions. *Australian & New Zealand Journal of Statistics*, 41(3):353–374, 1999.

S. Kullback. *Information theory and statistics.* John Wiley and Sons, New York, 1959.

S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

Alberto Luceño. Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimators. *Computational Statistics & Data Analysis*, 51(2):904–917, November 2006.

J. J. A. Moors. A quantile alternative for kurtosis. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(1):25–32, 1988.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

John S. Ramberg and Bruce W. Schmeiser. An approximate method for generating symmetric random variables. *Commun. ACM*, 15:987–990, 1972.

John S. Ramberg and Bruce W. Schmeiser. An approximate method for generating asymmetric random variables. *Commun. ACM*, 17(2):78–82, 1974.

John S. Ramberg, Pandu R. Tadikamalla, Edward J. Dudewicz, and Edward F. Mykytka. A probability distribution and its uses in fitting data. *Technometrics*, 21(2):201–214, May 1979.

Bo Ranneby. The maximum spacing method. An estimation method related to the maximum likelihood method. *Scandinavian Journal of Statistics*, 11(2):93–112, 1984.

David W. Scott. On optimal and data-based histograms. *Biometrika*, 66(3): 605–610, 1979.

Herbert A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, March 1926.

Steve Su. A discretized approach to flexibly fit generalized lambda distributions to data. *Journal of Modern Applied Statistical Methods*, 4(2): 408–424, 2005.

Steve Su. Fitting gld to data via quantile matching method. In Z. Karian and E. Dudewicz, editors, *Handbook of Fitting Statistical Distributions with R*. CRC Press/Taylor & Francis, 2010.

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

John W. Tukey. The future of data analysis. *The Annals of Mathematical Statistics*, 33(1):1–67, March 1962.