

# Estimating the Parameters of the Generalized Lambda Distribution

Asif Lakhany and Helmut Mausser

Fitting a probability distribution to observed or generated data constitutes an essential part of any data analysis system. The Generalized Lambda Distribution, while extremely versatile in this regard, is also a difficult distribution to fit. Parameter estimation methods that attempt to match moments or quantiles of the data require minimizing a bivariate non-linear function. The suitability of the resulting fitted distribution must be evaluated using a goodness-of-fit test and, if found unacceptable, the minimization procedure must be manually restarted from a new initial point. We propose a parameter estimation method that automatically generates initial points, as necessary, using quasi-random Sobol sequences.

Fitting a probability distribution to data is an important task in any statistical data analysis. The data to be modelled may consist of observed events, such as a financial time series, or it may comprise simulation results, such as a set of possible profit and loss values for a portfolio. In the former case the fitted distribution may form the basis for generating scenarios, while in the latter case it may be used for calculating risk measures, such as Value-at-Risk. When fitting data, one typically first selects a general class, or family, of distributions and then finds values for the distributional parameters that best match the observed data. Rachev and Mitnik (2000), for example, use generalized multivariate  $t$ -distributions and mixtures of distributions to accurately model the tails of market risk factors.

The Generalized Lambda Distribution (GLD), originally proposed by Ramberg and Schmeiser (1974), is a four-parameter generalization of Tukey's Lambda family (Hastings et al. 1947) that has proved useful in a number of different applications. Since it can assume a wide variety of shapes, the GLD offers risk managers great flexibility in modelling a broad range of financial data. Due to its versatility, however, obtaining

appropriate parameters for the GLD can be a challenging problem. An excellent synopsis of the GLD, its applications and parameter estimation methods appear in Karian and Dudewicz (2000).

The initial, and still the most popular, approach for estimating the GLD parameters is based on matching the first four moments of the empirical data. This is undoubtedly due in part to the availability of published tables that provide parameter values for given levels of skewness and kurtosis (see, e.g., Ramberg et al. (1979); Karian and Dudewicz (2000)). However, different parameter values can give rise to the same moments and so, while the tabulated parameters may match or closely approximate the first four moments, they may in fact fail to adequately represent the actual distribution of the data. Thus, as is well noted in the literature, a goodness-of-fit test should be performed to establish the validity of the results. If this test fails, or if the levels of skewness and kurtosis are outside of the tabulated values, it is necessary to use numerical procedures to find suitable parameters. Such procedures, which typically involve the downhill simplex method (Nelder

and Mead 1965) or some variant thereof, require as input an initial estimate of the parameters. If multiple local optima exist, the solution returned is contingent on this estimate. Thus, several attempts may be required before obtaining parameter values that are acceptable from a goodness-of-fit perspective.

Instead of matching moments, Öztürk and Dale (1985) minimize the total squared differences between the data and the expected values of order statistics implied by the G $\lambda$ D. The Nelder-Mead downhill simplex algorithm is used to find the optimal parameters. The method, which the authors call “least squares,” successfully fits a set of data for which tabulated moment-matching values are unavailable. As with moment matching, the resulting distribution must be assessed using a goodness-of-fit test, and several trials may be required before finding an acceptable solution.

Unlike previous approaches, King and MacGillivray (1999) assess the quality of the G $\lambda$ D directly by performing goodness-of-fit tests for specified combinations of parameter values. Their “starship” method can, however, be extremely slow if the search grid is poorly specified or if the sample size is large (since each goodness-of-fit test can take a significant amount of time in this case).

The problem of finding acceptable parameters when the initial goodness-of-fit test fails has received relatively little attention in the literature. Given knowledge of the desired distribution’s characteristics, such as its support (i.e., the range of values that are assigned positive probability) an analysis of the parameter space may provide guidance in selecting a starting point for the search procedure (see, for example, Karian and Dudewicz (2000)). In this paper, we show that controlled randomization of the initial point is a simple alternative that can provide good results with minimal intervention on the part of the user. Such “random restart” procedures are often incorporated in heuristic algorithms for solving difficult combinatorial optimization problems (see, for example, Glover (1986) and Maffioli (1986)). Essentially, we use the moment-matching or least squares methods to obtain a candidate set of parameters, conduct

a goodness-of-fit test and, if it fails, repeat the search starting from a new initial point. The initial points are generated using Sobol sequences (Sobol and Shukman 1993), which have the attractive property of more uniformly covering the sample space than points generated at random. We apply this approach to both the traditional parametrization of the G $\lambda$ D, due to Ramberg and Schmeiser (1974), as well as the parametrization of Freimer et al. (1988). To distinguish between them, we refer to these parametrizations as RS and FMKL, respectively.

This paper is organized as follows. First, we review the RS and FMKL parametrizations of the G $\lambda$ D and derive the moments of the latter. We then briefly describe the moment-matching, least squares and starship methods for fitting the G $\lambda$ D. Following this, we introduce the controlled randomization approach, and illustrate the procedure by fitting data from the literature (Öztürk and Dale 1985) and a sample from the financial area. Finally, we offer our conclusions and suggestions for further study.

### Parametrizations of the G $\lambda$ D

Tukey’s Lambda family of distributions is defined by the quantile function  $Q(u)$

$$Q(u) = \begin{cases} \frac{u - (1-u)^\lambda}{\lambda}, & \lambda \neq 0 \\ \frac{\log(u)}{1-u}, & \lambda = 0 \end{cases} \quad (1)$$

where  $0 \leq u \leq 1$ .

A four-parameter generalization of Equation 1, which we call the RS parametrization, was proposed in Ramberg and Schmeiser (1974). For this generalization, the quantile function is given by

$$Q(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2} \quad (2)$$

The properties of this distribution are studied in detail in Ramberg et al. (1979). In addition to elaborating the richness of this four-parameter G $\lambda$ D to fit a wide variety of frequency distributions, Ramberg et al. (1979) also note that the proposed distribution (Equation 2) is not defined for certain combinations of the parameters. A good summary of the shape

characteristics and the regions in which the distribution given by Equation 2 is well defined appears in King and MacGillivray (1999). The fact that the RS parametrization is not valid in some regions of the  $(\lambda_3, \lambda_4)$  plane for a given value of  $\lambda_2$  creates problems in implementing a parameter estimation algorithm, especially when such an algorithm relies on a search in the parameter space.

In order to avoid this problem, Freimer et al. (1988) devise a different parametrization for the GλD, denoted FMKL, which is given by

$$Q(u) = \lambda_1 + \frac{I}{\lambda_2} \left( \frac{u^{\lambda_3} - I}{\lambda_3} - \frac{(I-u)^{\lambda_4} - I}{\lambda_4} \right) \quad (3)$$

This parametrization is well defined over the entire two-dimensional plane for the parameters  $\lambda_3$  and  $\lambda_4$ . Note, however, that in order to have a finite  $k$ th order moment, it is necessary that

$$\min(\lambda_3, \lambda_4) > -\frac{I}{k}.$$

Although we shall not discuss the shape characteristics and asymptotic behaviour of the parametrizations in Equations 2 and 3 (this can be found in the references cited above), it is essential to understand the role played by each parameter in the GλD. To begin with,  $\lambda_1$  is the location parameter. The parameter  $\lambda_2$  determines the scale or the extent of the data. Finally, parameters  $\lambda_3$  and  $\lambda_4$  capture the shape characteristics of the empirical distribution generated by the data (for a symmetric distribution,  $\lambda_3 = \lambda_4$ ).

### Parameter estimation methods

Several methods for estimating the parameters of the GλD have been reported in the literature. In this section, we review briefly the moment-matching, least squares and starship methods.

#### Moment-matching method

The moment-matching method, described in this paper, was proposed in Ramberg and Schmeiser (1974). The method can be described in a straightforward manner as follows: given the GλD distribution with quantile function  $Q(u)$ , find parameters  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$  so that the mean  $\mu$ , variance  $\sigma^2$ , skewness  $\alpha_3$ , and kurtosis

$\alpha_4$  of the GλD match the corresponding mean  $\mu^*$ , variance  $(\sigma^*)^2$ , skewness  $\alpha_3^*$ , and kurtosis  $\alpha_4^*$  of the sample (i.e., the first four moments of the theoretical GλD match those of the data). More formally, if  $f(x) \equiv f(x; \lambda)$  denotes the probability density function of the random variable  $X$  with distribution  $F \equiv Q^{-1}$ , we compute the parameters  $\lambda$  such that

$$\mu \equiv E(X) \equiv \int_{-\infty}^{\infty} x f(x) dx = \mu^* \quad (4)$$

$$\sigma^2 \equiv E(X - E(X))^2 \equiv \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = (\sigma^*)^2 \quad (5)$$

$$\alpha_3 \equiv \frac{1}{\sigma^3} E(X - E(X))^3 \equiv \frac{1}{\sigma^3} \int_{-\infty}^{\infty} (x - \mu)^3 f(x) dx = \alpha_3^* \quad (6)$$

$$\alpha_4 \equiv \frac{1}{\sigma^4} E(X - E(X))^4 \equiv \frac{1}{\sigma^4} \int_{-\infty}^{\infty} (x - \mu)^4 f(x) dx = \alpha_4^* \quad (7)$$

In the following, we derive formulae that allow us to compute parameters  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$  satisfying Equations 4–7 to a desired level of accuracy (closed-form solutions do not exist). While approximation errors may have an impact on the quality of the fitted distribution to some degree, the fact remains that even if the first four moments are matched exactly, one cannot be assured that the resulting theoretical distribution will match perfectly the empirical distribution (Freimer et al. 1988). The quality of the fit can be ascertained only through a goodness-of-fit test.

Expressions of Equations 4–7 in terms of the parameters  $\lambda$  are readily available for the RS parametrization (see, for example, Ramberg and Schmeiser (1974)). For the sake of completeness, we derive these formulae for the FMKL parametrization.

Based on the identity  $F(Q(u)) = u$  and using the chain rule for differentiation, we obtain

$$f(x) = \frac{1}{\frac{\partial}{\partial u} Q(u)}$$

which means that if  $X$  is the random variable with quantile function  $Q(u)$ , then

$$E(X^k) = \int_0^1 (Q(u))^k du$$

Next, we note that for the FMKL parametrization (Equation 3) the quantile function can be rewritten as

$$F^{-1}(u) = \left( \lambda_1 - \frac{1}{\lambda_2 \lambda_3} + \frac{1}{\lambda_2 \lambda_4} + \frac{1}{\lambda_2} \left( \frac{\lambda_3}{\lambda_3} - \frac{(1-u)\lambda_4}{\lambda_4} \right) \right) \quad (8)$$

$$= (b + a\hat{Q}(u))$$

where

$$\hat{Q}(u) = \frac{\lambda_3}{\lambda_3} - \frac{(1-u)\lambda_4}{\lambda_4}$$

Now, if  $X$  represents the random variable with quantile function given by  $Q(u)$ , and  $Y$  denotes the random variable with quantile function  $\hat{Q}(u)$ , then in view of Equation 8 we have

$$E(X) = (aE(Y) + b)$$

$$E(X - E(X))^k = a^k E(Y - E(Y))^k$$

We now compute the following quantities:

$$v_k = \int_0^1 \left( \frac{u^{\lambda_3}}{\lambda_3} - \frac{(1-u)^{\lambda_4}}{\lambda_4} \right)^k du \quad (9)$$

Expanding the integrand on the right-hand side of Equation 9 using binomial expansion gives

$$v_k = \int_0^1 \sum_{j=0}^k \binom{k}{j} (-1)^j u^{\frac{\lambda_3(k-j)}{\lambda_3} - \frac{\lambda_4 j}{\lambda_4}} \frac{(1-u)^{\lambda_4 j}}{\lambda_4^j} du \quad (10)$$

$$= \sum_{j=0}^k \frac{(-1)^j}{\lambda_3^{k-j} \lambda_4^j} \binom{k}{j} \beta(\lambda_3(k-j) + 1, \lambda_4 j + 1)$$

The beta function on the right-hand side of Equation 10 is defined if both of its arguments are positive, which essentially means that the following holds:

$$\min_{k,j} \lambda_3(k-j) + 1 > 0$$

$$\min_j \lambda_4 j + 1 > 0 \quad (11)$$

From Equation 11 it is clear that the inequality is only crucial when  $\lambda_3, \lambda_4 < 0$ . Since  $0 \leq j \leq k$ , Equation 11 can be written as

$$\min(\lambda_3, \lambda_4) > \frac{-1}{k}$$

Using Equation 10 we obtain:

$$v_1 = \frac{1}{\lambda_3(\lambda_3 + 1)} - \frac{1}{\lambda_4(\lambda_4 + 1)}$$

$$v_2 = \frac{1}{\lambda_3^2(2\lambda_3 + 1)} + \frac{1}{\lambda_4^2(2\lambda_4 + 1)}$$

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

$$v_3 = \frac{1}{\lambda_3^3(3\lambda_3 + 1)} - \frac{1}{\lambda_4^3(3\lambda_4 + 1)}$$

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

$$+ \frac{3}{\lambda_3 \lambda_4^2} \beta(\lambda_3 + 1, 2\lambda_4 + 1)$$

$$v_4 = \frac{1}{\lambda_3^4(4\lambda_3 + 1)} + \frac{1}{\lambda_4^4(4\lambda_4 + 1)}$$

$$+ \frac{6}{\lambda_3^2 \lambda_4^2} \beta(2\lambda_3 + 1, 2\lambda_4 + 1)$$

$$- \frac{4}{\lambda_3^3 \lambda_4} \beta(3\lambda_3 + 1, \lambda_4 + 1)$$

$$- \frac{4}{\lambda_3 \lambda_4^3} \beta(\lambda_3 + 1, 3\lambda_4 + 1)$$

Denoting the  $k$ th central moment of the distribution  $F^{-1}(u)$  by  $\mu_k$ , we obtain

$$\mu_2 = \frac{1}{\lambda_2^2} (v_2 - v_1^2)$$

$$\mu_3 = \frac{1}{\lambda_2^3} (v_3 - 3v_1 v_2 + 2v_1^3)$$

$$\mu_4 = \frac{1}{\lambda_2^4} (v_4 - 4v_1 v_3 + 6v_1^2 v_2 - 3v_1^4)$$

and, accordingly, the standardized skewness and kurtosis are given by:

$$\alpha_3 = \frac{v_3 - 3v_1 v_2 + 2v_1^3}{(v_2 - v_1^2)^{\frac{3}{2}}} = G_3(\lambda_3, \lambda_4)$$

$$\alpha_4 = \frac{v_4 - 4v_1 v_3 + 6v_1^2 v_2 - 3v_1^4}{(v_2 - v_1^2)^2} = G_4(\lambda_3, \lambda_4)$$

Thus, the values of the parameters  $\lambda_3$  and  $\lambda_4$  may be computed by solving the following system of nonlinear equations:

$$\begin{aligned} G_3(\lambda_3, \lambda_4) &= \alpha_3^* \\ G_4(\lambda_3, \lambda_4) &= \alpha_4^* \end{aligned}$$

in the region  $(-1/4, \infty) \times (-1/4, \infty)$  of the  $(\lambda_3, \lambda_4)$  plane or equivalently, minimizing the quantity

$$H(\lambda_3, \lambda_4) = \|G - \alpha\|_p \tag{12}$$

in some suitable Euclidean  $p$ -norm (in our experiments we have used the 2-norm). In Equation 12,  $G = [G_3, G_4]^T$  and  $\alpha = [\alpha_3^*, \alpha_4^*]^T$ .

Finally, once the values for  $\lambda_3$  and  $\lambda_4$  are obtained, the remaining parameters are computed using the formulae:

$$\lambda_2 = \frac{\sqrt{v_2 - v_1^2}}{\sigma^*} \tag{13}$$

$$\lambda_1 = \mu^* + \frac{1}{\lambda_2} \left( \frac{1}{\lambda_3 + 1} - \frac{1}{\lambda_4 + 1} \right) \tag{14}$$

**Least squares method**

The least squares method for computing the parameters of the GλD (Öztürk and Dale 1985) can be described as follows. Let  $x_i, i = 1, \dots, n$  denote the  $i$ th order statistic of the data which is to be represented by the quantile function  $Q(u)$  (for the sake of definiteness we use the FMKL representation of  $Q(u)$ ) and let  $U_i, i = 1, \dots, n$  denote the order statistic of the corresponding uniformly distributed random variable  $F(X)$ . The least squares method finds the values of  $\lambda$  for which the differences between the observed and predicted order statistics are as small as possible. Thus, it minimizes the function

$$G(\lambda) = \sum_{i=1}^n \left( x_{(i)} - \lambda_1 - \frac{Z_i}{\lambda_2} \right)^2 \tag{15}$$

where

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

The formulae for computing  $EU_{(i)}^{\lambda_3}$  and  $E(1 - U_{(i)})^{\lambda_4}$  are (Mykytka and Ramberg 1979)

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

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

Note that the function to be minimized in Equation 15 depends on all four parameters. To avoid solving a computationally demanding minimization problem in four-dimensional space,  $(\lambda_1, \lambda_2)$  are decoupled from  $(\lambda_3, \lambda_4)$  as in the moment-matching method. We first assume that  $\lambda_3$  and  $\lambda_4$  are constant and solve the minimization problem for  $\lambda_1$  and  $\lambda_2$ . The values for  $\lambda_1$  and  $\lambda_2$ , once obtained, are substituted into the minimization function and then compute  $\lambda_3$  and  $\lambda_4$ . We adopt this strategy based on the observation that the expression for  $G(\lambda)$  has  $\lambda_1$  and  $\lambda_2$  in linear form, whereas  $\lambda_3$  and  $\lambda_4$  appear in nonlinear form. Thus, differentiating  $G(\lambda)$  with respect to  $\lambda_1$  and  $\lambda_2$  and setting the resulting function equal to zero, we obtain

$$\begin{aligned} \lambda_1 &= \mu_x - b_{xZ} \mu_Z \\ \lambda_2 &= \frac{1}{b_{xZ}} \end{aligned} \tag{16}$$

where  $\mu_x$  and  $\mu_Z$  denote the mean of sample data and the quantities  $Z_i$ , respectively, and the regression coefficient  $b_{xZ}$  is given by

$$b_{xZ} = \frac{i=1}{n} \frac{\sum_{i=1}^n (x_{(i)} - \mu_x)(Z_i - \mu_Z)}{\sum_{i=1}^n (Z_i - \mu_Z)^2} \tag{17}$$

Inserting Equations 16 and 17 in Equation 15, we obtain, after some rearrangement of terms,

$$G(\lambda_3, \lambda_4) = (1 - r_{xZ}(\lambda_3, \lambda_4))^2 \sum_{i=1}^n (x_{(i)} - \mu_x)^2$$

where  $r_{xZ}$  is the correlation coefficient between the quantities  $x_i$  and  $Z_i$ . Thus, in order to

minimize  $G(\lambda_3, \lambda_4)$  we need to maximize the quantity  $r_{xZ}(\lambda_3, \lambda_4)$  or, equivalently, minimize the function

$$H(\lambda_3, \lambda_4) = -r_{xZ}(\lambda_3, \lambda_4)^2 \quad (18)$$

Once  $\lambda_3$  and  $\lambda_4$  have been obtained by minimizing Equation 18, they can be inserted into Equations 16 and 17 in order to compute  $\lambda_1$  and  $\lambda_2$ .

Before leaving this section, we again emphasize that maximizing the square of the correlation coefficient  $r_{xZ}$  does not guarantee that the estimated parameters will yield a distribution that closely matches the empirical distribution of the sample data.

### Starship method

The starship method proposed by King and MacGillivray (1999) directly addresses the issue of fit quality in the sense that it evaluates candidate parameter values by means of a goodness-of-fit test. In its simplest form, the method can be described as follows:

**Step I:** Select a region in four-dimensional space that covers the range of the four parameters  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_3$  and  $\lambda_4$  appropriately.

**Step II:** On the region selected in Step I, overlay a four-dimensional rectangular grid.

**Step III:** Evaluate the grid points created in Step II by performing a goodness-of-fit test on the corresponding distributions. Once the test is satisfied, stop; otherwise, continue with the next point in the grid (alternatively, one can examine all grid points and select the one with the best goodness-of-fit measure).

Thus, instead of minimizing a (two-dimensional) function of  $\lambda_3$  and  $\lambda_4$  in order to match certain characteristics of the sample data, the starship method conducts a four-dimensional grid search. There is only one stopping criterion, which is based on the outcome of a goodness-of-fit test. Although the starship method is very flexible, covers a wider range for the parameters and is guaranteed to fit well (because of the goodness-of-fit test included in the algorithm), it is also time consuming as the search is performed in

four-dimensional, rather than two-dimensional, space. Note that it is possible to reduce this time somewhat by carefully selecting the range over which the four-dimensional grid is created and then adaptively remeshing the grid based on the previous results (King and MacGillivray 1999).

### Randomized restart method

As we have noted, parameters that replicate certain characteristics of the data do not necessarily produce a distribution that adequately fits the sample. Thus, while the moment-matching and least squares methods can locate candidate parameter values relatively quickly, by solving a two-dimensional optimization problem, the resulting distributions should be verified by a goodness-of-fit test. If the test fails, it is necessary to manually restart the optimization from a new initial solution. While the starship method directly evaluates fit quality, performing a large number of goodness-of-fit tests becomes computationally impractical for large sample sizes. A procedure that automates the restart of algorithms such as moment matching and least squares provides a method for fitting the GLD that is not only self-contained, but also computationally efficient. We now propose one such procedure that constructs quasi-random initial solutions using Sobol sequences to obtain uniform coverage of the solution space. Before formalizing the procedure, we describe two of its main components: the Nelder-Mead downhill simplex method and the Kolmogoroff-Smirnoff goodness-of-fit test.

### Downhill simplex method

Both the moment-matching and least squares methods require minimizing a function of  $\lambda_3$  and  $\lambda_4$ , which we have denoted by  $H(\lambda_3, \lambda_4)$  for the sake of uniformity, albeit making note of the fact that the actual function is different for each method. In both cases, the complexity of  $H(\lambda_3, \lambda_4)$  favours a solution method that requires only function evaluations, rather than derivatives. Thus, in a manner consistent with Ramberg et al. (1979) and Öztürk and Dale (1985), we use the downhill simplex method of Nelder and Mead (1965) to minimize  $H(\lambda_3, \lambda_4)$ .

Details for implementing the downhill simplex algorithm can be obtained from any standard book on numerical methods in optimization (see, for example, Gill et al. (1982)). The algorithm, as applied to minimizing a function of  $n$  variables, can be briefly described as follows. Starting with a non-degenerate simplex, evaluate the function at each of the  $n + 1$  vertices. The worst vertex ( $V_W$ ), which has the highest function value, is reflected about the hyperplane generated by the remaining  $n$  vertices in order to obtain a new vertex ( $V_N$ ). The function is evaluated on this newly generated vertex and if  $V_N$  is also the new worst vertex then a contraction of the simplex is performed along the  $\overrightarrow{V_W V_N}$  direction, while if  $V_N$  is the best vertex then an expansion of the simplex is performed along the  $\overrightarrow{V_W V_N}$  direction. This is illustrated in Figure 1, where the reflection to  $P_1^{(r)}$  is extended to arrive at the new point  $P_1^{(e)}$ . This simplex-generation process continues until a stopping criterion is reached. This could, for example, consist of the change in function value meeting some minimum tolerance from one iteration to the next, or some maximum number of iterations being reached. Various extensions of this algorithm are discussed in Press et al. (1992).

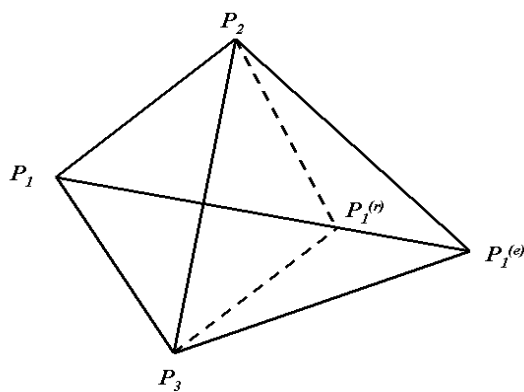


Figure 1: Simplex expansion

**Kolmogoroff-Smirnoff test**

The Kolmogoroff-Smirnoff (KS) test (see, for example, D’Agostino and Stephens (1986)), which tests a set of data for consistency with a specified distribution, proceeds as follows. Let  $x_{(i)}$

denote the  $i$ th order statistic of a given sample of size  $n$ . We construct the empirical cumulative distribution function  $S_n(x_{(i)}) = \frac{i}{n}$  and, given a set of parameter values  $\lambda_1, \lambda_2, \lambda_3$  and  $\lambda_4$ , a theoretical cumulative distribution function  $F(x)$ . We then compute the KS statistic:

$$D_n = \max_i |S_n(x_{(i)}) - F(x_{(i)})|$$

The null hypothesis, which states that the sample is drawn from  $F$ , is rejected if  $D_n$  exceeds a specified critical value (e.g., at the 95% confidence level, the null hypothesis can be rejected if  $D_n > \frac{1.36}{\sqrt{n}}$ ).

In our case, a potential concern when using the KS test relates to the support of the theoretical distribution, denoted  $supp(F)$ . If  $F$  has finite support (which is quite possible in the case of GLD) then some of the data points  $x_{(i)}$  may fall outside  $supp(F)$ . To deal with this possibility, we assume that  $F(x) = 1$  for all points  $x$  that lie above  $supp(F)$ , and that  $F(x) = 0$  for all points  $x$  that lie below  $supp(F)$ . This represents a conservative approach as it increases the possibility of obtaining a large  $D_n$  value and, therefore, rejects the null hypothesis. In our implementation, we also allow for placing a limit on the number of points that are excluded from the support. Parameters resulting in distributions that exclude more than the allowed number of points are considered unacceptable.

**Procedural description**

We iteratively invoke the downhill simplex algorithm, starting from a Sobol-generated initial simplex, until a distribution is obtained that satisfies the KS test at the 95% confidence level:

**Step I:** Partition the search region  $R = \cup R_i$ , composed of (possibly disconnected) regions  $R_i$  where the function  $H(\lambda_3, \lambda_4)$  is defined, into  $M$  subregions (again possibly disconnected)  $S_i$ . To each region  $S_i$  associate a subjective probability  $p_i$ , which can be viewed as the probability of finding the distribution fit in that particular region. At this stage it is possible to exploit a priori knowledge about the distribution of the data that one may have—which is likely to be the

case in many practical situations—and also the information about the shape of the distribution in those regions that are well documented in the literature. For example, a U-shaped distribution can be avoided by setting the probability at zero in the region where  $1 \leq \lambda_3 \leq 2$  and  $1 \leq \lambda_4 \leq 2$ , more compactly denoted  $[1, 2] \times [1, 2]$ . Note, however, that these regions affect only the starting point of the search; we do not explicitly prevent the algorithm from subsequently entering a region, other than those in which  $H(\lambda_3, \lambda_4)$  is not defined.

**Step II:** Based on the initial probabilities (these can be modified dynamically, if necessary) we create a random number generator  $G_N$  that generates numbers between 1 and  $M$  with corresponding probabilities  $p_i, i = 1, \dots, M$ .

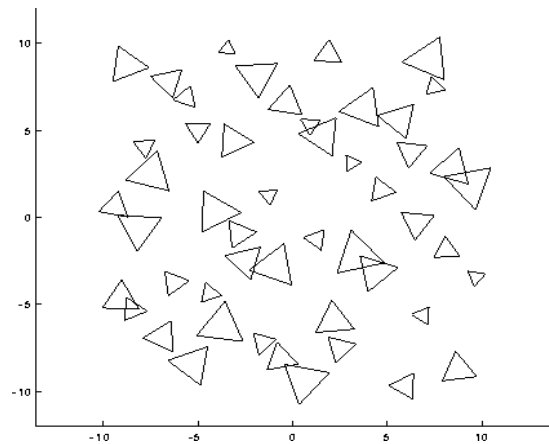
**Step III:** Create a Sobol sequence generator,  $G_S$ , in four dimensions (see, for example, Sobol and Shukman (1993) and Shukman (1994) for details on Sobol sequences). The first two dimensions can be mapped to a location in each region after suitable transformation based on the geometry of the region. The third and the fourth coordinates control the size and the orientation of the simplex, respectively. The upper and lower limits for the diameter of the simplex can be either user controlled or automatically computed as some fraction of the smallest dimension of the region concerned. Note that Sobol sequences have the property of filling regions uniformly. This is well illustrated in Figure 2 for 50 Sobol simplices within the region  $[-10, 10] \times [-10, 10]$ .

**Step IV:** Sample a number  $n$  from the generator  $G_N$  and an appropriately scaled vector  $(\tilde{\lambda}_3, \tilde{\lambda}_4, r_S, \theta_S)$  from  $G_S$ . Select region  $S_n$  and generate a simplex with centroid located at  $(\tilde{\lambda}_3, \tilde{\lambda}_4)$ , with diameter  $2r_S$  and orientation  $\theta_S$  (note that we are generating equilateral triangles and thus exploiting the symmetry) in  $S_n$ . Perform the downhill simplex search and arrive at the minimal value for the function  $H$  at coordinates  $(\lambda_3, \lambda_4)$ . Using these values, compute the remaining parameters using Equations 13 and 14 in the case of the moment-matching method or

Equation 16 in the case of the least squares method.

**Step V:** The theoretical G $\lambda$ D defined by the computed parameters in the previous test is subjected to a Kolmogoroff-Smirnoff test in order to verify that the fitted distribution matches the empirical distribution. If the goodness-of-fit test fails, then the process is repeated starting from Step IV. Otherwise, the iterative simplex search is terminated.

Note that the above procedure inherently assumes that lower values of  $H(\lambda_3, \lambda_4)$  correspond to better goodness-of-fit measures.



**Figure 2:** Sobol simplices

## Numerical experiments

To validate our implementation of the moment-matching, least squares and starship algorithms, we apply them to a sample data set from Öztürk and Dale (1985). The data set, containing 75 observations, has sample moments  $\mu^* = 5.109$ ,  $\sigma^* = 1.013$ ,  $\alpha_3^* = 1.009$  and  $\alpha_4^* = 3.344$ .

These values of skewness and kurtosis are outside the range of tabulated values, and so fitting must be performed with numerical procedures in this case.

We consider the downhill simplex algorithm to have converged when the relative difference between the function values at the three simplex vertices, defined as the difference between the highest and lowest values divided by the average value, is less than 0.001. A candidate solution is

	RS parametrization				FMKL parametrization		
	Öztürk and Dale (1985)	Moment matching	Least squares	Starship	Moment matching	Least squares	Starship
$\lambda_1$	6.0217	3.8433	5.9794	4.3320	6.7216	6.9965	6.5000
$\lambda_2$	0.4525	0.1780	0.4547	0.2344	0.0728	0.0820	0.0650
$\lambda_3$	8.2956	0.0000	8.2163	0.0500	4.0263	3.5012	4.4000
$\lambda_4$	0.9201	0.2908	0.9830	0.2000	11.2648	13.8369	11.0000
Number of searches	N/A	1	5	3	1	21	274
Time (s)	N/A	0.02	0.07	0.01	0.01	0.25	0.97
Adjusted KS	0.5259	0.4596	0.5612	1.2341	0.9419	0.9171	1.3410

**Table 1:** Results for data from Öztürk and Dale (1985)

accepted only if the resulting distribution satisfies the following conditions:

- all of the data points are contained in its support
- it has an adjusted KS statistic ( $D_n\sqrt{n}$ ) less than 1.36 (this corresponds to a confidence level of 95% for the KS test).

The starship method uses 10 grid points in each dimension, for a total of 10,000 possible sets of parameter values. For the RS parametrization, the initial grid spans the range  $[4.332, 5.659] \times [0.2344, 1.3034] \times [0, 0.50] \times [0, 0.50]$ , which was obtained by examining the characteristics of the data as described in King and MacGillivray (1998). For the FMKL parametrization, we simply use the range  $[6.5, 7.0] \times [0.05, 0.10] \times [3.50, 4.50] \times [11, 14]$ , based on the results of the moment-matching and least squares approaches.

Table 1 compares the results of Öztürk and Dale with several other approaches (note that the columns denoted “Moment matching” and “Least squares” refer to the automated restart procedure described in the previous section). The solution found by Öztürk and Dale is very close to that obtained by our implementation of least squares, although we required five restarts

before finding a  $G\lambda D$  satisfying the former two conditions. The FMKL parametrization is generally more difficult to fit than the RS one, both in terms of computational complexity and solution quality (as reflected by the relative sizes of the KS statistics). We found that, for this example, the FMKL parametrization often gave distributions whose support excluded one or more observations. It is interesting to note that moment matching yielded a valid solution on the first iteration for both parametrizations.

Next, we consider a larger set of financial data and fit the  $G\lambda D$  to the value distribution of a portfolio of callable bonds, obtained from a simulation of 1,000 Monte Carlo scenarios. The data has sample moments  $\mu^* = 33.886$ ,  $\sigma^* = 0.010$ ,  $\alpha_3^* = -0.527$  and  $\alpha_4^* = 2.985$ , which allows the parameters to be obtained from existing moment-matching tables for the RS parametrization. In this case, we allow, at most, one observation to be excluded from the support of the fitted distribution. The starship method uses an initial grid on  $[33.8794, 33.8936] \times [19.1611, 44.9191] \times [0, 0.50] \times [0, 0.50]$  for the RS parametrization and  $[33.88, 33.89] \times [129.7, 129.8] \times [0, 0.50] \times [0, 0.50]$  for the FMKL parametrization.

	RS parametrization				FMKL parametrization		
	Moments (Table)	Moment matching	Least squares	Starship	Moment matching	Least squares	Starship
$\lambda_1$	33.8934	33.8932	33.8928	33.8908	33.8873	33.8874	33.8870
$\lambda_2$	19.7535	19.9238	19.5667	24.3127	129.7466	127.6109	129.7000
$\lambda_3$	0.2440	0.2423	0.2315	0.2500	0.1113	0.1208	0.0500
$\lambda_4$	0.0551	0.0585	0.0603	0.1000	0.3444	0.3601	0.2500
Number of searches	N/A	4	6	3127	316	10	1897
Time (s)	N/A	0.22	1.30	88.66	14.12	1.56	89.38
Adjusted KS	0.7430	0.7215	0.7291	1.3208	1.0495	1.0126	1.2629

**Table 2:** Results for callable bond portfolio

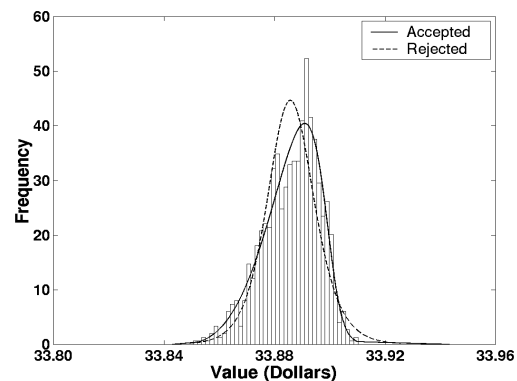
Table 2 shows the results of fitting a GLD to the empirical value distribution. Again, the RS parametrization tends to be preferable in terms of fit quality and computation time. The starship method is clearly inefficient for this larger data set; although we accept the first solution with an adjusted KS statistic less than 1.36, the starship method still performs significantly more goodness-of-fit tests than the other approaches.

The moment-matching and least squares methods all require multiple restarts before finding an acceptable fit. This is particularly evident when applying moment matching to the FMKL parametrization. Further investigation showed that 315 solutions were rejected because the resulting distribution's support excluded more than one of the observed data points.

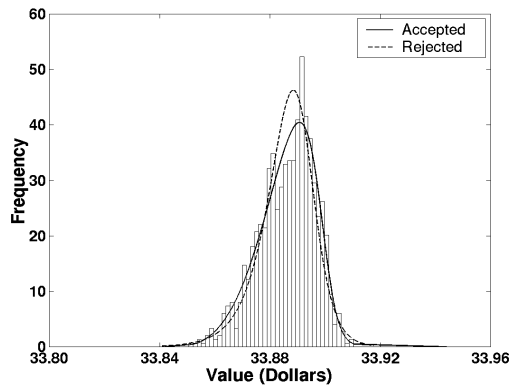
For both RS and FMKL parametrizations, we found it necessary to exclude otherwise acceptable distributions because of an insufficiently large support. Thus, the ability to perform multiple searches can be particularly useful when it is necessary to obtain a good fit in the tails of the distribution, as is the case when calculating VaR, for example.

Figures 3 and 4 illustrate the differences between accepted and rejected solutions for the moment-

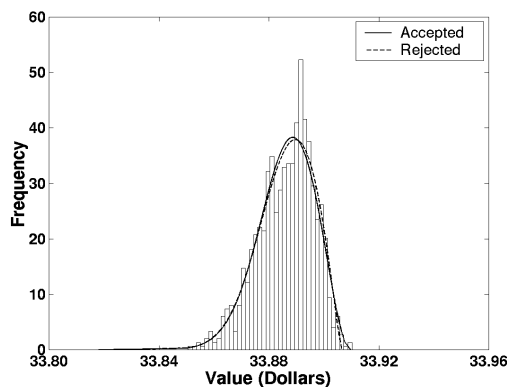
matching and least squares methods, respectively, under the RS parametrization. In both of these cases, the rejected solutions fail the goodness-of-fit test. In contrast, Figures 5 and 6 show, for the FMKL parametrization, rejected solutions that satisfy the goodness-of-fit test, but fail to adequately span the range of observed data. The random restart procedure is an effective way of obtaining solutions that meet multiple objectives, such as an acceptable fit quality and an adequate support, that might be imposed by the user.



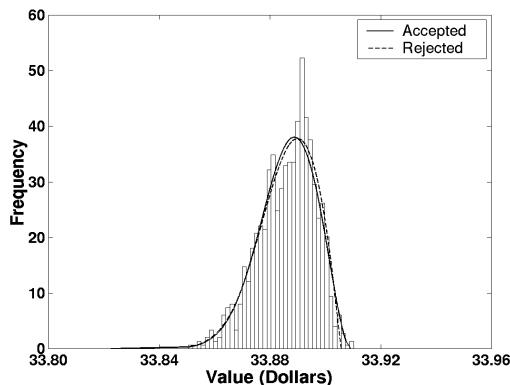
**Figure 3:** Fit of RS parametrization using moment matching



**Figure 4:** Fit of RS parametrization using least squares



**Figure 5:** Fit of FMKL parametrization using moment matching



**Figure 6:** Fit of FMKL parametrization using least squares

## Conclusions

When fitting the GLD to data, it is important that the distributions found by methods such as moment matching or least squares be properly evaluated. As is evident in our examples, such distributions may have a support that does not adequately span the observed data, or they may fail to satisfy a goodness-of-fit test. Thus, it is often necessary to perform multiple searches in the parameter space before finding an acceptable distribution. We have proposed a procedure that automatically restarts the downhill simplex algorithm by generating a quasi-random initial simplex at each iteration. Sobol sequences are used to ensure uniform coverage of the parameter space. The procedure is computationally efficient; it performs fewer goodness-of-fit tests than the starship method, for example, which makes it attractive when dealing with large samples.

Our generation of the initial simplex is independent of the search process itself—we attempt to locate the simplices more or less uniformly throughout the parameter space, regardless of which regions may have been previously traversed by the downhill simplex method. Using the knowledge gained during previous searches to locate a better initial simplex, or even to guide the downhill simplex algorithm itself, represents an interesting extension to this approach. Concepts from tabu search (Glover and Laguna 1997) may prove to be effective in this regard.

## References

- D'Agostino, R. and M. Stephens, 1986, "Goodness-of-fit techniques," v. 68, *Statistics: Textbooks and Monographs*, New York: Marcel Dekker.
- Freimer, M., G. Mudholkar, G. Kollia and C. Lin, 1988, "A study of the generalized Tukey Lambda family," *Communications in Statistics, Theory and Methods*, 17(10): 3547–3567.
- Gill, P., M. Wright, and W. Murray, 1982, *Practical Optimization*, California: Academic Press.
- Glover, F., 1986, "Future paths for integer programming and links to artificial intelligence," *Computers and Operations*

- Research*, 13(5): 533–549.
- Glover, F. and M. Laguna, 1997, *Tabu Search*, The Netherlands: Kluwer Academic Publishers.
- Hastings, C., F. Mosteller, J. Tukey and C. Winsor, 1947, “Low moments for small samples: a comparative study of statistics,” *Annals of Mathematical Statistics*, 18: 413–426.
- Karian, Z. and E. Dudewicz, 2000, *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*, Boca Raton: CRC Press.
- King, R. and H. MacGillivray, 1999, “A Starship estimation method for the generalized lambda distributions,” *Australian and New Zealand Journal of Statistics* 41(3): 353–374.
- Maffioli, F., 1986, “Randomized algorithms in combinatorial optimization: a survey,” *Discrete Applied Mathematics*, 14: 157–170.
- Mykytka, E. and J. Ramberg, 1979, “Fitting a distribution to data using an alternative to moments,” *IEEE Proceedings of the 1979 Winter Simulation Conference*, 361–374.
- Nelder, J. and R. Mead, 1965, “A simplex method for function minimization,” *Computer Journal*, 7: 308–313.
- Öztürk, A. and R. Dale, 1985, “Least squares estimation of the parameters of the Generalized Lambda Distribution,” *Technometrics*, 27(1): 81–84.
- Press, W., S. Teukolsky, W. Vetterling and B. Flannery, 1992, *Numerical Recipes in C*, 2nd ed., Cambridge: Cambridge University Press.
- Rachev, S. and S. Mittnik, 2000, *Stable Paretian Models in Finance*, New York, NY: Wiley & Sons.
- Ramberg, J., E. Dudewicz, P. Tadikamalla and E. Mykytka, 1979, “A probability distribution and its uses in fitting data,” *Technometrics*, 21(2): 201–214.
- Ramberg, J. and B. Schmeiser, 1974, “An approximate method for generating asymmetric random variables,” *Communications of the ACM*, 17(2): 78–82.
- Shukhman, B., 1994, “Generation of quasi-random ( $LP_\tau$ ) vectors for parallel computation,” *Computer Physics Communications*, 78: 279–286.
- Sobol, I. and B. Shukman, 1993, “Random and quasi random sequences: numerical estimates of uniformity of distribution,” *Mathematical and Computer Modelling*, 18(8): 39–45.