Title: Estimation of the generalized lambda distribution parameters for grouped data

Key Words: Distribution fitting, quantile function, goodness of fit.

Abstract: In this paper we consider the problem of fitting a five-parameter generalization of the Lambda distribution to data given in the form of a grouped frequency table. The estimation of parameters is done by six different procedures: percentiles, moments, probability-weighted moments, minimum Cramér-Von Mises, maximum likelihood, and pseudo least squares. These methods are evaluated and compared using a Monte Carlo study where the parent populations were GLD approximations of Normal, Beta, Gamma random variables and for nine combinations of sample sizes and number of classes. Of the estimators analyzed is concluded that, although the method of pseudo least squares suffers from a number of limitations, it appears to be the candidate procedure to estimate the parameters of a GLD from grouped data.

1. Introduction

The generalized lambda distribution (GLD) is a flexible and manageable tool for modeling empirical and theoretical distributions. In empirical work the data on continuous variables are generally presented in one of two ways: individual observations are reported or the data are summarized in a grouped form with the frequency associated with each group being reported as it is frequently the case for size distribution. The estimation of the parameters of the GLD for continuous data has been discussed in several papers (*e.g.* Lakhany and Mausser, 2000) and there are several methods to estimate the parameters when grouped data only are available (*e.g.* Berkson, 1980). Things are more difficult in the case of the GLD because the density and the distribution function are not available directly except in a few special cases and must be determined by numerical methods. This paper analyzes and compares traditional and new methods of parameter estimation from grouped data for a quantile function. More specifically, our purpose is to determine the method of estimation that preserves the versatility and the effectiveness of the GLD in fitting continuous variables when

i) the data are collected or published in a histogram form or in the form of a grouped frequency table. Individual observations are unrecoverable or permanently lost;

ii) the grouping is not fine enough to be negligible.

The GLD appears to be most useful when the form of the empirical distribution is found to vary considerably and it seems likely that a model with four or five parameters would be needed to describe the data.

The content of the present paper is organized as follows: in Section 2 the properties of a five-parameter GLD model are described and the main analytical and statistical peculiarities are summarized. Section 3 contains a discussion of six estimation procedures in the case of grouped data paying special attention to the extension of these methods to a random variable defined by its quantile function. A limited series of Monte Carlo experiments is undertaken in Section 4 with the objective of identifying the most effective estimator for the parameters of a GLD model.

2. The generalized lambda distribution (GLD)

In this section we shall outline the defining properties of a five-parameter generalization of the lambda distribution in order to obtain smoothed analytic representations for grouped data. The GLD is specified by the quantile function

$$X(p,\lambda) = \lambda_1 + \lambda_2 p^{\lambda_4} - \lambda_3 q^{\lambda_5} \qquad 0 \le p \le 1,\ q = 1-p \tag{1}$$

where λ_1 is a location parameter, λ_2, λ_3 are linear parameters prevalently related to the scale of the variable and λ_4, λ_5 are exponential parameters determining the shape of the quantile function. The two linear parameters contribute to specify the relative weights of the tails and to avoid imposing upon the fitted curves constraints not present in the data. Since the GLD features four shape parameters, we can expect this model to assume a wide variety of shapes.

The expression (1) readily yields X(p,λ) in terms of a uniform random variable p on the interval [0,1]. This fact is particularly relevant not only for simulation experiment, but also for order statistics, optimal grouping, inequality measures, heavy tail behavior analysis, loss distributions, osculatory interpolation, Q-Q plotting.

The following conditions are imposed:

$$\begin{array}{llll} \textit{If } \lambda_2 \to 0 \textit{ then } \lambda_4 > -\infty; & \textit{If } \lambda_4 \to \infty \textit{ then } |\lambda_2| < \infty; & \textit{If } \lambda_4 \to -\infty \textit{ then } \lambda_2 \neq 0 \\ \textit{If } \lambda_3 \to 0 \textit{ then } \lambda_5 > -\infty; & \textit{If } \lambda_5 \to \infty \textit{ then } |\lambda_3| < \infty; & \textit{If } \lambda_5 \to -\infty \textit{ then } \lambda_3 \neq 0 \end{array} \tag{2}$$

The support of the GLD random variable is $(\lambda_1 - \lambda_3, \lambda_1 + \lambda_2)$ if $\lambda_4, \lambda_5 > 0$ and is the real line if $\lambda_4, \lambda_5 < 0$. Hence, the extremes of X(p, λ) are finite or infinite according to the sign of the exponential parameter.

Analytic expression for the cumulative distribution function F(x,λ) of a GLD model is in general not available. However, the fact that the GLD is not invertible is not a serious drawback because the same is true for many popular models such as normal, lognormal,

generalized gamma, generalized beta. The limiting forms as λ_4 and/or λ_5 tend to zero or diverge to ∞ are

GLD (λ_1, λ_2, λ_3, ∞, ∞) = GLD (λ_1, 0, 0, λ_4, λ_5) = degenerate
GLD (λ_1, λ_2, 0, λ_4, λ_5) = GLD (λ_1, λ_2, $\lambda_3<\infty$, λ_4,∞) = power function
GLD (λ_1, 0, λ_3, $\lambda_4>-\infty$, λ_5) = GLD (λ_1, $\lambda_2<\infty$, λ_3, ∞, λ_5) = generalized Pareto
GLD (λ_1, λ_2, λ_3, 1, 1)= uniform ($\lambda_1-\lambda_3$, $\lambda_1+\lambda_2$)
GLD (λ_1, $\lambda_2=\lambda_3$, 2, 2)= uniform ($\lambda_1-\lambda_2$, $\lambda_1+\lambda_2$)

Other versions of the uniform distribution are present. The GLD has been found to adapt, by suitable choice of λ, to a wide variety of theoretical and practical distributions (Karian and Dudewicvz, 2000). For example, if we let $\lambda_2\lambda_4=\beta$, $\lambda_3\lambda_5=\gamma$, $\lambda_1=\alpha-\beta/\lambda_4+\gamma/\lambda_5$ then

GLD (α, $\beta=\gamma$, 0, 0) = logistic
GLD (α, 0, γ, λ_4, 0) = exponential
GLD (α, β, 0, 0, λ_5) = reflected exponential

There are several parameterization of the asymmetric lambda distribution: Shapiro-Wilk (1965), Joiner-Rosenblatt (1971), Ramberg-Schmeiser (1974), Filliben (1975), Hoaglin (1975), Lam *et al.* (1980), Freimer *et al.* (1988), Sarabia (1996), Devroye (1996), Gilchrist (2000). All of these versions can be obtained as special cases of (1).

Although the GLD is very versatile and flexible in curve fitting, the estimation of its parameters can be difficult. In particular, the large number of parameters can generate high variability of the estimates, strong correlation between those estimates, numerical singularities due to heavy cancellation in $X(p,\lambda)$, and time consuming optimization algorithms. Of course, if the histogram of the data or theoretical considerations suggest fitting simpler models (*e.g.* exponential, Pareto or generalized Pareto or a symmetric distribution) then the more parsimonious parameterization should be employed.

Shapes of the GLD model

The probability density function of a GLD random variable is defined implicitly by the density quantile function (d.q.f), that is the density expressed in terms of p

$$\frac{1}{\frac{dX(p;\lambda)}{dp}} = h\left[X(p;\lambda)\right] = \frac{1}{\lambda_4\lambda_2 p^{\lambda_4-1} + \lambda_5\lambda_3 q^{\lambda_5-1}} \tag{3}$$

See Parzen (1979) for a general definition of this function. The parameter λ_1 exerts its influence on the density function via the quantile function X(p, λ).

If $\lambda_2=\lambda_3$ and $\lambda_4=\lambda_5$ then the GLD is symmetric about the pole $X=\lambda_1$ because the quantile function satisfies the condition $X(p, \lambda)- \lambda_1=-X(1-p,\lambda)+ \lambda_1$. Hence interchanging ($\lambda_2$, λ_4) and (λ_3, λ_5) in (3) we obtain a d.q.f. that is the mirror image of the original d.q.f. When scale and location are changed we transform the variable Y=a+bX. The transformed distribution is another member of the GLD family with λ_1, λ_2, λ_3 replaced by $a+b\lambda_1$, $b\lambda_2$, $b\lambda_3$ respectively. Expression h[X(p, λ)] represents a legitimate probability density function if and only if it is nonnegative and integrates to one. The latter condition follows directly from (3). Regions in which h[X(p, λ)] is non negative are

$$V_1 : \lambda_4\lambda_2 > 0,\ \lambda_5\lambda_3 \geq 0 \qquad V_2 : \lambda_4\lambda_2 \geq 0,\ \lambda_5\lambda_3 > 0$$

$$V_3 : \operatorname{sgn}(\lambda_4\lambda_2) \neq \operatorname{sgn}(\lambda_3\lambda_5), \lambda_4 < 1, \lambda_5 > 1 \qquad V_4 : \operatorname{sgn}(\lambda_4\lambda_2) \neq \operatorname{sgn}(\lambda_3\lambda_5), \lambda_4 > 1, \lambda_5 < 1$$

provided that *provided that*

$$-\frac{\lambda_4\lambda_2}{\lambda_5\lambda_3} > \frac{(1-\lambda_4)^{1-\lambda_4}(\lambda_5-1)^{\lambda_5-1}}{(\lambda_5-\lambda_4)^{\lambda_5-\lambda_4}} \qquad -\frac{\lambda_3\lambda_5}{\lambda_2\lambda_4} > \frac{(1-\lambda_5)^{1-\lambda_5}(\lambda_4-1)^{\lambda_4-1}}{(\lambda_4-\lambda_5)^{\lambda_4-\lambda_5}} \tag{4}$$

In particular, regions V_3 and V_4 are based on a result of Karian, Dudewicz, McDonald (1996). The ordinates of the density quantile function at the extremes of the range of variation are ($1/\lambda_5\lambda_3$, $1/\lambda_4\lambda_2$) if λ_4, $\lambda_5 \geq 1$ and zero for λ_4, $\lambda_5 < 1$.

The parameters λ_4 and λ_5 determine the type of tails of the GLD. For example, if λ_4, $\lambda_5 > 0$ then (3) has increasingly peakedness and short tails; if λ_4, $\lambda_5 < 0$ the tails have increasingly heaviness.

The density tends to zero both as $p \to 0$ and as $p \to 1$ if, respectively, $\lambda_4 < 1$ and $\lambda_5 < 1$, provided that λ_2 and λ_3 are finite. On the other hand, if $\lambda_4 \geq 1$ ($\lambda_5 \geq 1$) the density has truncated left (right) tail. In particular, the case $\lambda_4 > 1$ allows model (1) to describe the actual distribution with a range equal to the closed-open infinite interval $[\lambda_1 - \lambda_3/\lambda_4, \infty]$ which is physically reasonable for many applications. The relative weights of the tails are controlled via λ_2 and λ_3.

The derivative of the quantile-density function (3) is

$$h'\left[X(p,\lambda)\right] = \left\{h\left[X(p,\lambda)\right]\right\}^3 \left[\lambda_3\lambda_5(\lambda_5 - 1)q^{\lambda_5 - 2} - \lambda_2\lambda_4(\lambda_4 - 1)p^{\lambda_4 - 2}\right] \tag{5}$$

If $p \to 0$ or if $p \to 1$ then $h'[X(p, \lambda)]$ tends, respectively, to

$$\begin{aligned} &-\frac{(\lambda_4 - 1)}{\operatorname{sgn}(\lambda_2\lambda_4)(\lambda_2\lambda_4)^2} p^{-2\lambda_4 + 1} \quad for \;\; p \to 0 \;\; if \quad \lambda_4 < 1 \\ &\frac{(\lambda_5 - 1)}{\operatorname{sgn}(\lambda_3\lambda_5)(\lambda_3\lambda_5)^2} q^{-2\lambda_5 + 1} \quad for \;\; p \to 1 \;\; if \quad \lambda_5 < 1 \end{aligned} \tag{6}$$

Therefore, the slope at the lower (upper) extreme of a GLD random variables is zero, nonzero but finite, or infinite if, respectively, $\lambda_4 < 0.5$, $\lambda_4 = 0.5$, $0.5 < \lambda_4 < 1$ ($\lambda_5 < 0.5$, $\lambda_5 = 0.5$, $0.5 < \lambda_5 < 1$). If $1 < \lambda_4 < 2$ then $h'[X(p, \lambda)]$ diverges to $-\infty$ when $\lambda_2 > 0$ (or $+\infty$ when $\lambda_2 < 0$) for $p \to 0$. If $1 < \lambda_5 < 2$ then $h'[X(p, \lambda)]$ diverges to $+\infty$ when $\lambda_3 > 0$ (or $-\infty$ when $\lambda_3 < 0$) for $p \to 1$. The derivative (5) is always finite at p=0 and p=1 if λ_4, $\lambda_5 \geq 2$.

If p_c is such that $h'[X(p_c, \lambda)] = 0$ then

$$h''\left[X(p_c;\lambda)\right] = -h^3\left[X(p_c;\lambda)\right]\left\{\lambda_2\lambda_4(\lambda_4 - 1)(\lambda_4 - 2)p_c^{\lambda_4 - 3} + \lambda_3\lambda_5(\lambda_5 - 1)(\lambda_5 - 2)q_c^{\lambda_5 - 3}\right\} \tag{7}$$

Hence, the density is unimodal if λ_4, λ_5>2 and λ_2, λ_3>0 or if λ_4, λ_5<0 and λ_2, λ_3<0 or if 0<λ_4, λ_5<1 and λ_2, λ_3>0. The density is zeromodal if 0<Min{λ_4, λ_5} and Max{λ_4 λ_5}>1 and λ_2, λ_3>0 or if λ_4=1,2 and sgn($\lambda_3\lambda_5$)>0 or if λ_5=1,2 and sgn($\lambda_2\lambda_4$)>0.

The moments of the GLD random variable

The mean and the median of a GLD random variable are

$$\mu = \lambda_1 + \frac{\lambda_2}{(\lambda_4+1)} - \frac{\lambda_3}{(\lambda_5+1)}; \quad M_e = \lambda_1 + \lambda_2 0.5^{\lambda_4} - \lambda_3 0.5^{\lambda_5} \tag{8}$$

Consider the linear transformation Z=X-λ_1. The expected value of Z^i is given by

$$\mu_i(\lambda) = \sum_{j=0}^{i} \binom{i}{j} (-1)^j \lambda_2^{i-j} \lambda_3^j B(\lambda_4(i-j)+1, \lambda_5 j+1) \tag{9}$$

where B(x,y) denotes the complete beta function. Since Z-E(Z)=X-E(X) the central moments of X coincide with the central moments of Z (Freimer *et al.*, 1988). More specifically

$$\begin{aligned}
\sigma^2 &= \alpha_1 - 2cB(\lambda_4+1, \lambda_5+1) \\
\gamma_1 &= \frac{\alpha_2}{\sigma^3} - \frac{3c^2}{\sigma^3}\left[\frac{B(2\lambda_4+1, \lambda_5+1)}{\lambda_3} - \frac{B(\lambda_4+1, 2\lambda_5+1)}{\lambda_2}\right] \\
\gamma_2 &= \frac{\alpha_3}{\sigma^4} - \frac{4c^3}{\sigma^4}\left[\frac{B(3\lambda_4+1, \lambda_5+1)}{\lambda_3^2} + \frac{B(\lambda_4+1, 3\lambda_5+1)}{\lambda_2^2} - \frac{3B(2\lambda_4+1, 2\lambda_5+1)}{2c}\right] \\
\alpha_i &= \frac{\lambda_2^{i+1}}{[(i+1)\lambda_4+1]} + (-1)^{i+1}\frac{\lambda_3^{i+1}}{[(i+1)\lambda_5+1]}, i=1,2,3; \quad c=\lambda_2\lambda_3
\end{aligned} \tag{10}$$

It is evident that the i-th moment of the GLD exists if and only if min(λ_4,λ_5)>1. In addition, $1+\gamma_1^2 \le \gamma_2$ as for all random variables. In spite of its versatility, the GLD cannot be fitted to all data sets because there are moment values that no GLD random variable can achieve. For example, symmetric distributions with a kurtosis value of γ_2<1.75 (Joiner and

Rosenblatt, 1971). The coefficient of skewness $\gamma_1(\lambda)$ is function of four parameters which makes it difficult to establish what shape changes occur as two or more parameters vary simultaneously (see, for example, MacGillivray, 1982). Moreover, for each value of $\text{sgn}[\gamma_1(\lambda)]$ the shape of a GLD can range from normal curve to decidedly non-normal curve either positively or negatively skewed. Nevertheless, ever, the degree of skewness can also be measured by a more stable index

$$\frac{\mu - M_e}{S_{Me}} = b(\lambda) = \frac{\frac{\lambda_2}{(\lambda_4+1)}\left[1-(\lambda_4+1)0.5^{\lambda_4}\right] - \frac{\lambda_3}{(\lambda_5+1)}\left[1-(\lambda_5+1)0.5^{\lambda_5}\right]}{\frac{\lambda_2}{(\lambda_4+1)}\left[1-0.5^{\lambda_4}\right] + \frac{\lambda_3}{(\lambda_5+1)}\left[1-0.5^{\lambda_5}\right]} \qquad (11)$$

where S_{Me} is the mean deviation about the median. Index $b(\lambda)$, suggested by Bonferroni (1940-41, p. 79, p. 116) and Groeneveld (1986), varies between -1 and 1 with values near -1 indicating a truncated J-shaped density and near 1 for a truncated L-shaped density.

For the GLD, interchanging λ_2 and λ_3, other things being fixed, reverses the type of the skewness. In fact, $b(\lambda_2, \lambda_3, \lambda_4, \lambda_5) = -b(\lambda_3, \lambda_2, \lambda_4, \lambda_5)$; also, $b(\lambda_2,-\lambda_3, \lambda_4, \lambda_5)=b(-\lambda_2, \lambda_3,\lambda_4, \lambda_5)$. For the special case of $\lambda_2=\lambda_3$ the coefficient (11) depends only on (λ_4, λ_5). If the distribution is symmetric (that is, $\lambda_2=\lambda_3$, $\lambda_4=\lambda_5$) then $b(\lambda)=0$. The converse is not necessarily true unless $\lambda_4=\lambda_5$. In fact, if $\lambda_2\neq\lambda_3$ then $b(\lambda)$ vanishes if and only if $\lambda_4=\lambda_5=1$, but for this parameter combination the GLD becomes the uniform distribution which is symmetric.

3. Parameter estimation

In this section we show how to determine unknown parameters when the GLD is fitted to empirical distribution reported in the form of a frequency table. Suppose that a sample of size n from a GLD is grouped into k intervals

$$\left(X_{i-1}, X_i\right],\ n_i,\quad N_i = \sum_{j=1}^{i} n_j;\quad p_i = \sum_{j=1}^{i} f_j,\qquad f_j = \frac{n_j}{n}\qquad i = 1,2,\ldots,k \tag{12}$$

The values $\{X_i,\ i=0,1,2,\ \ldots,\ k;\ X_{i-1}<X_i\}$ are the boundaries of k exhaustive and non-overlapping classes and $k\geq 3$ is limited in practice (Kariya, 1986). In the present paper the number of classes is fixed in advance and the X_i's are known constants and we will consider the problem of estimation of the parameter λ when the available information consists only of the number of observed values X_i falling into the various classes.The probability mass assigned to the i-th class can be written as

$$F\left(X_i;\lambda\right) - F\left(X_{i-1};\lambda\right) = \pi_i(\lambda) > 0,\ i = 1,2,\ldots,k;\quad \sum_{i=1}^{k} \pi_i(\lambda) = 1 \tag{13}$$

We suppose that $\lambda \in \Lambda \supset R^5$ and we think of $\lambda_0 \in \Lambda$ as the true but unknown value of the parameter λ. Clearly, the grouping scheme may significantly affect the parameter estimation and the variance of estimators (*e.g.* O'Neill and Wells, 1972). For instance, if the observations cluster significantly around particular values producing multimodal distributions, no GLD can give an acceptable agreement with this behavior.

Percentile matching estimates (PM)

This method consists of equating a selection of five empirical X_i and five theoretical percentiles $X(p_i,\lambda)$ provided that X_i is a good approximation of the unknown empirical order statistic corresponding to p_i.

$$X_{i_j} = \lambda_1 + \lambda_2 p_{i_j}^{\lambda_4} - \lambda_3 q_{i_j}^{\lambda_5};\quad j = 1,\ldots,5 \tag{14}$$

Where $i_j \in (1,2,\ldots,k')$ and k'=k if the upper bound of the variable is known and finite otherwise k'=k-1. The PM method has the advantage of being operative without the necessity of knowing every measurement (for example, the endpoints of the terminal classes can be ignored). On the other hand, the PM estimates depend markedly on the particular choice of percentage points. If the selected percentiles favor the central part of the distribution then the estimation is better for the values around the mode, but at the cost of being much worse in the tails. If more constraints are placed on the tails then fewer conditions can be imposed on remaining percentiles. In addition, the use of too extreme percentiles would lead to very inaccurate estimates because of the high variance of the corresponding order statistics. The grouping effect increases in the central classes because of the increased crowding of the order statistics (David and Mishriky, 1968). If k'>5 is not too large then all the ${}_{k'}C_5$ combinations of the k' observed percentiles taken five at a time could be investigated to establish the "optimal" subset of percentiles (supposing that at least one set gives an admissible value for λ). Alternatively, one may select (systematically or at random) a prespecified number of five-percentile solutions. These values are then combined in a suitable way to produce more efficient estimates (see for example Castillo and Hady, 1996). No distribution theory seems to be available for this procedure.

To keep the computation at a reasonable level, the PM estimates were obtained applying system (14) to the five sextiles w_1, w_2, w_3, w_4, w_5 computed by

$$w_s = (1-\beta_s)X_{j-1} + \beta_s X_j; \quad \beta_s = \frac{p_s - p_{j-1}}{f_j}, \quad p_s = \frac{s}{6}; \quad j = \underset{0<i\le k}{Min}\{p_i \ge p_s\}, \quad s = 1,2,\ldots,5 \qquad (15)$$

We have considered other methods for interpolating quantiles (Schmeiser and Deutsch,1977; Harrell and Davis, 1982; Korn *et al.* 1997) but did not offer any improvement over the linear interpolation. It is important to note that, as five percentiles of the GLD tend to coincide with the corresponding observed percentiles, the remaining empirical percentiles do not necessarily concentrate more and more about their theoretical true value. Since (14) is a nonlinear system of equations in λ, one needs to use an iterative procedure to

solve for λ. To calculate the PM estimates of λ we have applied the downhill simplex minimization to the following criterion

$$S_{QM}(\lambda) = \underset{1\le s\le 5}{Max}\left\{\left|w_s - X(p_s, \lambda)\right|\right\} \tag{16}$$

under the constraint that $\min(\lambda_4,\lambda_5)>-1$. This method is a well-known derivative-free optimization algorithm due to Nelder and Mead (1965). It requires only function evaluations and has a wide applicability for general function minimization (Olsson and Nelson, 1975). A limitation of the direct-search approach is the lack of guarantee that the global optimum will be achieved, though it works very well on a range of practical problems. Note that there may be more than one value for which the minimum of the criterion is attained (Karian and Dudewicz, 1999) and that there is no assurance that the algorithm will terminate in the interior of the appropriate parameter space. In many cases, a typical data set will lead to parameter estimates satisfying their natural constraints (4). For other cases the constraints must be explicitly imposed. To limit the evaluations of the objective function in the required region, the direct-search algorithm handled the constraints by setting the objective function equal to 10^{42} whenever an inequality was violated.

Moment matching estimates (MM)

This method suggests that empirical moments should be found from observed data and the GLD model employed which has the closest theoretical counterparts. The empirical moments are calculated using the class midpoints

$$m_1 = \sum_{j=1}^{k} c_j f_j = m_1; \qquad m_i = \sum_{j=1}^{k} \left[c_j - m_1\right]^i f_j, \;\; i = 2,\ldots,5 \tag{17}$$

The grouping errors introduced by the use of the class midpoints in skewed distributions can be significant though, as the greatest class width goes to zero while the number of observations is limited, the bias tends to zero.

The computation of the moment estimates involves the solution of

$$\mu_i(\lambda) = m_i, \quad i=1,\ldots,5 \tag{18}$$

Under the restriction $\min(\lambda_4,\lambda_5)>-0.20$. A set of roots of (18) can be obtained by applying the downhill simplex by to the objective function defined as

$$S_{MM}(\lambda) = \sqrt{\sum_{i=2}^{5}\left(\mu_i(\lambda) - m_i\right)^2} \tag{19}$$

Several other objective function can be used to find MM estimates, but expression (19) has proved to be the most convenient criterion function for the particular problem we are solving. The presumed value of λ_1 is quickly computed from (8) once the other parameters have been estimated. It should be remarked that (18) can have multiple solutions or no real solution for some data sets (Ramberg *et al.*, 1980). Even when a solution exists, the numerical procedure devoted to its search may miss it because of convergence failure (Ramberg *et al*., 1979). However, Karian *et al*. (1996) state that the possibility of multiple solutions to (18) it is not a problem, but can be an opportunity.

The classical method of moments is restricted to distributions possessing fairly light tails because they must have a finite fifth moment. Moreover, the sample moments are sensitive to extreme observations or other contamination in data and sampling variability in high order moments can be very large. If the interval lengths are unequal, then the estimated moments admit of no simple "Sheppard's-like" corrections. Thus, using sample estimates of μ_3, μ_4, μ_5 in a fitting procedure may lead to extremely biased estimates. Another possible drawback is that, even though the discrepancy between empirical moments and theoretical moments is extremely low, there is no guarantee that the corresponding model fits adequately the data. Also, the method of moments assumes the availability of specific measurements for each class; consequently, if X_0 and/or X_k are omitted, then the MM estimates will be biased as long as their imputation or the centering of the

extreme classes does not typify adequately the tails of the distribution. Often the method of moments fails because of the lack of information on the larger values.

Probability-weighted moment estimates (PWM)

The method of probability-weighted moments proposed by Greenwood *et al.* (1979) provides a viable alternative to the traditional method of moments when the tails of the empirical distribution indicate that high order moments of the population may not be finite. In fact, PW moments can be defined for any random variable whose mean exists in the finite sense. For the GLD it is convenient to work with the probability-weighted moments

$$\tau_i = E\left[Q(p;\lambda)p^i\right] = \frac{\lambda_1}{i+1} + \frac{\lambda_2}{\lambda_4 + i + 1} - \frac{\lambda_3 i!}{\prod_{j=0}^{i}(\lambda_5 + j + 1)} \quad i = 0,1,\ldots,4 \tag{20}$$

Expression (20) shows that this type of moments is particularly of interest for a random variable known in terms of its quantile function. Sample probability-weighted moments for grouped data can be given as follows

$$t_0 = \frac{\sum_{j=1}^{n} c_j}{n}; \qquad t_i = \frac{\sum_{j=i+1}^{n} c_j \left\lfloor \prod_{r=1}^{i}(j-r) \right\rfloor}{\prod_{r=0}^{i}(n-r)}, \quad for \;\; i = 1,2,3,4 \tag{21}$$

In practice, the empirical PWM's for grouped data are computed on the basis of no grouping effect in the data collection and using class midpoints as "observed values" with a GLD density. Thus n_1c_1's, n_2c_2's, …, n_kc_k's constitute the sample to which the PWM estimators have been applied.

In the method of PW moments, the vector λ is estimated by computing the first five sample PW moments (21), setting them equal to the population PW moments (20) and solving the system of equations

$$\tau_i(\lambda) = t_i, \quad i = 1, \ldots, 4 \tag{22}$$

for $\lambda_2, \lambda_3, \lambda_4, \lambda_5$. An estimate for λ_1 is then obtained by means of (8). In general, the PWM estimates for λ do not lend themselves into the explicit forms. To determine the PWM estimates we proceed analogously to the method of moments. More specifically, the down-hill simplex algorithm is applied to the criterion

$$S_{PWM}(\lambda) = \max_{2 \le s \le 5} \left\{ \left| t_i - \tau_i(\lambda) \right| \right\} \tag{23}$$

subject to $\min(\lambda_4, \lambda_5) > -1$. The main advantage of using PWM is that the parameter estimates are more reliable than the MM estimates for heavy-tails distributions because the observed values appear linearly in (21). Furthermore, the equations associated with PWM estimators are simpler and the computational techniques devoted to their solution experience less convergence difficulties than the traditional moment estimators. A possible drawback is that the sampling and asymptotic properties of the PWM estimators for grouped data are not yet well established.

Minimum Cramér-von Mises estimates (MCM)

The quality of fit can be measured by the discrepancy between the empirical distribution function and the predicted cumulative frequencies thought to be an element of the GLD family. In particular, we can use the minimum distance estimation method based on a grouped-data version of the Cramér-von Mises statistic (see Duchesne *et al.* 1997). The corresponding estimator is chosen to be that value of λ that minimizes

$$S_{MCM}(\lambda) = \sum_{i=1}^{k} \left[p_i - F(X_i, \lambda) \right]^2, \; F(X_i, \lambda) > F(X_{i-1}, \lambda), \quad F(X_0, \lambda) = 0 \tag{24}$$

where $F(X_i, \lambda)$ denotes the estimated p value that would generate the observation X_i for the given λ. Expression (24) weights each deviation equally; other interests may suggest a

different weighting scheme. If $S_{MCM}(\lambda)$ attains its minimum at an interior point $\lambda \in \Lambda$ then λ is also a solution of

$$\frac{\partial S_{MCM}(\lambda)}{\partial \lambda_r} = -\sum_{i=1}^{k'} \left[p_i - F(X_i, \lambda) \right] \frac{\partial F(X, \lambda)}{\partial \lambda_r} = 0 \qquad r = 1, 2, \ldots, 5 \tag{25}$$

According to Eubank (1998) we have

$$\frac{\partial F(X,\lambda)}{\partial \lambda_r} = -h\left[X(F_i,\lambda) \right] X_r'(F_i,\lambda) \qquad with \quad X_r'(F_i,\lambda) = \left[\frac{\partial X(p,\lambda)}{\partial \lambda_r} \right]_{p=F_i} \tag{26}$$

where $h[X(p,\lambda)]$ is the density-quantile function and $F_i = F(X_i,\lambda)$. To find λ we can use the Newton-Raphson iterative procedure. The elements of the gradient and Hessian can be written as

$$g_r(\lambda) = \sum_{i=1}^{k'} (p_i - F_i) h\left[X(F_i,\lambda) \right] X_r'(F_i,\lambda) \qquad r = 1, 2, \ldots, 5 \tag{27}$$

$$\begin{aligned} W_{r,s}(\lambda) = \sum_{i=1}^{k'} h^2\left[X(F_i,\lambda) \right] \Big\{ & X_s'(F_i,\lambda) X_r'(F_i,\lambda) \left[1 - h\left[X(F_i,\lambda) \right] A(F_i,\lambda)(p_i - F_i) \right] + \\ & - (p_i - F_i) X_s'(F_i,\lambda) B_r(F_i,\lambda) \Big\} \end{aligned} \tag{28}$$

Where

$$\begin{aligned} A(F_i,\lambda) &= \lambda_3 \lambda_5 (\lambda_5 - 1)(1 - F_i)^{\lambda_5 - 2} - \lambda_2 \lambda_4 (\lambda_4 - 1) F_i^{\lambda_4 - 2} \\ B_r(F_i,\lambda) &= \left[\frac{dX_r'(p,\lambda)}{dp} \right]_{p=F_i}; \qquad r = 1, 2, \ldots, 5 \end{aligned} \tag{29}$$

Given a preliminary estimate $\lambda^{\mathbf{0}}$ for λ then an algorithm to obtain the MCM estimates is described by the recursion

$$\lambda^{m+1} = \lambda^m - B_m \delta(\lambda^m); \qquad with \quad \delta(\lambda^m) = \left[W(\lambda^m) + \gamma_m I \right]^{-1} g(\lambda^m) \tag{30}$$

where B_m is a diagonal matrix of correction factors which fixes the step length of each parameter and γ_m is a positive scalar large enough to make $[W(\lambda^m)+\gamma_m I]$ positive definite when $W(\lambda^m)$ it is not. It should be remarked that the parameter estimates are highly correlated and have relatively large variances because of the extreme susceptibility of the GLD to even small changes in λ. The nonzero elements of B_m can be selected by performing a systematic search along the Newton direction by means of Faure quasi-random sequences (Faure, 1982; Fox, 1986). More precisely, we have considered a Faure sequence of 1000 points in the region $[-\delta(\lambda^m),\delta(\lambda^m)]$. The computation of F_i cannot be directly carried out because the cumulative distribution of the GLD model is not analytically invertible. To solve $X_i=X(F_i,\lambda)$ for $F_i>F_{i-1}$ and $F_0=0$ we have used a combination of bisection and Newton's method (see Press *et al.*, 1993, p. 366-367)

The estimators that minimize (24) belong to the family of minimum distance methods (Parr, 1981). These methods of estimation, under modest regularity conditions, are strongly consistent, asymptotically normal and perform well in a variety of settings. Furthermore, The procedure outlined above is close to the starship technique discussed by King and MacGillivray (1999).

Maximum likelihood estimates (ML)

The class frequencies can be thought of as a random sample of size n from a multinomial distribution specified by a vector $[\pi_1(\lambda),\pi_2(\lambda),\ldots,\pi_k(\lambda)]$ of probabilities of k mutually exclusive events. The negative log-likelihood function is of the form

$$S_{ML}(\lambda) = -\sum_{i=1}^{k} n_i Log\left[F(X_i;\lambda) - F(X_{i-1};\lambda)\right] = -\sum_{i=1}^{k} n_i Log\left[\pi_i(\lambda)\right], \qquad \sum_{i=1}^{k} n_i = n \qquad (31)$$

This can be minimized over λ through application of the scoring method (Rao, 1973, 366-374; McDonald and Ransom, 1979). The efficient score for the parameter λ_r and the (r,s)-term of the information matrix are

$$g_r(\lambda) = \sum_{i=1}^{k} \frac{n_i}{\pi_i(\lambda)} \left[\frac{\partial \pi_i(\lambda)}{\partial \lambda_r} \right]$$
$$W_{r,s}(\lambda) = \sum_{i=1}^{k} \frac{n}{\pi_i(\lambda)} \left[\frac{\partial \pi_i(\lambda)}{\partial \lambda_r} \right] \left[\frac{\partial \pi_i(\lambda)}{\partial \lambda_s} \right] \quad (32)$$

$$\frac{\partial \pi_i(\lambda)}{\partial \lambda_r} = -\left[h\left[X(F_i, \lambda) \right] X_r'(F_i, \lambda) - h\left[X(F_{i-1}, \lambda) \right] X_r'(F_{i-1}, \lambda) \right] \quad (33)$$

The method of scoring is very similar to the Newton-Raphson procedure and we can again use the recursion (30) to achieve the minimization of (31). Gilchrist (2000, p. 295-298) describes a more direct approach to the ML estimation for the quantile function based on individual observations. By standard theory, the ML estimators will be asymptotically normal, unbiased, consistent and efficient. See also Cheng and Iles (1987).

Pseudo least squares estimates (PLS)

The vector λ can be estimated taking up a nonlinear regression framework

$$X_i = E(X_i) + e_i, \qquad i = 1, 2, \ldots, k' \quad (34)$$

Since we ignore if some observation is equal to X_i, the value of X_i is not necessarily the order statistics corresponding to the observed p_i and $E(X_i)$ may not be the value predicted by (34). Yet, for a sample size sufficiently large and if the grouping is not too coarse, X_i can be considered a reasonable approximation to $X(p_i, \lambda)$ (David and Mishriky, 1968). It should be emphasized that classical assumptions on the error terms $\{e_i\}$ are inappropriate because the $\{X_i\}$ will not have equal variance nor will they be uncorrelated or come from a symmetric distribution. Since our purpose is to obtain an approximate solution, we ignore these violations. The expected value of order statistics from a GLD is available in closed form and we may write (34) as

$$X(p_i, \lambda) = \lambda_1 + \lambda_2 \left[\frac{\Gamma(\lambda_4 + N_i)\Gamma(n+1)}{\Gamma(N_i)\Gamma(\lambda_4 + n + 1)} \right] - \lambda_3 \left[\frac{\Gamma(\lambda_5 + n + 1 - N_i)\Gamma(n+1)}{\Gamma(n+1-N_i)\Gamma(\lambda_5 + n + 1)} \right] \quad (35)$$

provided that $\min(\lambda_4,\lambda_5)>-1$. The computation of $E(X_i)$ requires repeated evaluations of the gamma function. Nevertheless, it is well known that for continuous random variable, $E(X_i)$ converges to $X(p_i,\lambda)$. Thus, for large samples, we have

$$X_i = \lambda_1 + \lambda_2 p_i^{\lambda_4} - \lambda_3 q_i^{\lambda_5} + e_i; \quad i = 1,\ldots,k' \tag{36}$$

The least squares approach calls for choosing λ to minimize

$$S_{PLS}(\lambda) = \sum_{i=1}^{k} \left[X_i - X(p_i,\lambda)\right]^2 f_i \tag{37}$$

The $\{f_i\}$ in (37) take into account the fact that there are unequal counts in the intervals. The distinguishing features of many real data sets are a heavy thickness of the tails and an accentuate peakedness. It would be interesting to incorporate in the PLS method an ingenious system of weights so that the tails become adequately detectable.

Since $(\lambda_1,\lambda_2,\lambda_3)$ are in linear form, minimization of $S_{PLS}(\lambda)$ can be achieved by first assuming that (λ_4,λ_5) are known constants and then solving the minimization problem as a linear regression of X_i on $X(p_i,\lambda)$ (Lawton and Sylvestre, 1971). The parameter reduction has the additional benefit that it is necessary to supply an initial guess value only for (λ_4,λ_5). The values of $(\lambda_1,\lambda_2,\lambda_3)$ are then inserted into the minimization function (37) and a new value of (λ_4,λ_5) is computed. The process is repeated until an optimum of the criterion is reached. The current estimate of (λ_4,λ_5) can be determined using the Newton-Raphson algorithm described in the preceding section.

The criterion $S_{PLS}(\lambda)$ is the squared Samuel-Bachi distance (Samuel and Bachi, 1964) between the observed quantile function and the GLD. If X_i were the order statistic corresponding to p_i then the vector λ which minimizes $S_{PLS}(\lambda)$ would coincide with the weighted L^2 quantile distance estimator discussed by LaRiccia (1982). Oztürk and Dale (1985) used (26) for microdata. Oztürk and Abouammoh (1987) considered the absolute version of (26) for estimating λ in the Ramberg-Schmeiser parameterization of the GLD.

4. Comparison of the methods

A good fitting procedure is not very useful unless one can assess its standard error by Monte Carlo methods. This, however, is possible only when the procedure is completely automated and its execution time is foreseeable and compatible with an adequate number of replications. Unfortunately, the estimates for grouped data must be developed on a case-by-case basis because the results are not invariant under different grouping schemes even though the individual observations are the same. Consequently, only a limited simulation study was conducted to compare the performances of the alternative estimation techniques and investigate their properties.

Monte Carlo experiments

Random samples were obtained from GLD models which were good substitutes of two symmetric distributions: Normal(0,1), Beta(2,2) and two positively skewed distributions: Gamma(2,1), Beta(2,4).

Model	*Density*	*range*
Normal(0,1)	$(2\pi)^{-0.5}\ e^{-x^2}$	$-\infty,\ +\infty$
Beta(2,2)	$6x(1-x)$	0, 1
Gamma(2,1)	xe^{-x}	0, ∞
Beta(2,4)	$20x(1-x)^3$	0, 1

(38)

We believe that the above models are illustrative of a range of commonly encountered situations.

Parameters of the GLD models were selected by considering the closeness of the quantile function X(p,λ) to $F^{-1}()$ where F() is the cumulative distribution function of the random variable being approximated. To check the closeness we have followed the same idea as Karian and Dudewicz (2000, p.66-67) and used the criterion

$$d(GLD,F) = \underset{1\le i\le 499}{Max}\left|h(p_i,\lambda) - f\left[X(p_i,\lambda)\right]\right|; \quad p_i = \frac{i}{500},\ i = 1,2,\ldots,499 \tag{39}$$

Where f() is the probability density function corresponding to F. Parameter values used in simulations are reported in the following table.

Table 1: experimental cases from the GLD model

Model	λ_1	λ_2	λ_3	λ_4	λ_5	d(GLD,F)
Normal(0,1)	0.0000000	4.6525421	4.6525421	0.1492010	0.1492010	0.0004430
Beta(2,2)	0.5000000	0.5095575	0.5095575	0.4480095	0.4480095	0.0060918
Gamma(2,1)	24.0345788	1.0875235	24.0726325	0.4043437	0.0547165	0.0022480
Beta(2,4)	0.7617832	0.2707079	0.7662556	0.4462439	0.2448268	0.0059850

Simulations were performed for samples of size n∈(1000, 2000, 4000) classified into k classes where k∈ (8, 14, 20). All in all, we have examined 36 distinct configurations. The boundaries of the grouping intervals are as follows

Model	*class boundaries*
Normal(0,1)	$X_i = -3.1 + 6.2\left(\frac{i}{k}\right),\ X_{n-i} = 3.1 - 6.2\left(\frac{i}{k}\right) \quad i = 0,1,2,\ldots,k/2$
Beta(2, 2)	$X_i = \left(\frac{i}{k}\right) \quad i = 0,1,2,\ldots,k$
Gamma(2,1)	$X_i = \begin{cases} k=8 \Rightarrow & 0.0,\ 0.3,\ 0.7,\ 1.1,\ 1.9,\ 2.9,\ 4.2,\ 5.6,\ 8.4 \\ k=14 \Rightarrow & 0.0,\ 0.2,\ 0.4,\ 0.6,\ 0.8,\ 1.2,\ 1.6,\ 2.0,\ 2.4 \\ & 3.0,\ 3.4,\ 3.9,\ 4.8,\ 5.6,\ 8.4 \\ k=20 \Rightarrow & 0.0,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9 \\ & 1.0,\ 1.2,\ 1.4,\ 1.6,\ 2.0,\ 2.4,\ 2.8,\ 3.4,\ 4.0 \\ & 4.8,\ 5.6,\ 8.4 \end{cases}$
Beta(2, 4)	$X_i = \begin{cases} k=8 \Rightarrow & 0.00,\ 0.08,\ 0.16,\ 0.24,\ 0.32,\ 0.45,\ 0.60,\ 0.75,\ 1.00 \\ k=14 \Rightarrow & 0.00,\ 0.04,\ 0.08,\ 0.12,\ 0.16,\ 0.22,\ 0.28,\ 0.34,\ 0.40 \\ & 0.45,\ 0.50,\ 0.58,\ 0.67,\ 0.75,\ 1.0 \\ k=20 \Rightarrow & 0.00,\ 0.04,\ 0.08,\ 0.12,\ 0.16,\ 0.20,\ 0.24,\ 0.28,\ 0.32 \\ & 0.36,\ 0.40,\ 0.44,\ 0.48,\ 0.52,\ 0.56,\ 0.60,\ 0.64.\ 0.68 \\ & 0.74,\ 0.80,\ 1.00 \end{cases}$

(40)

Theoretically, Normal(0,1) and Gamma(2,1) have an infinite range. In practice, truncation at one or both the extreme is necessary as a result of the measurement mechanism or from physical considerations or by experimental convenience. Accordingly, we deal only with

truncated forms of these distributions by eliminating values lower than X_0 or greater than X_k (this implies the additional condition that $\lambda_4, \lambda_5 \geq 0$). Furthermore, the simulated distributions were discarded when they presented one or more zero observed frequencies.

For each GLD model the various estimates of λ were compared to the true value λ_0 Two simple coefficients of performance have been considered for comparison: the mean relative bias and the standard deviation

$$Mrb(\lambda_j) = \frac{1}{N}\left(\sum_{i=1}^{N}\left|\frac{\lambda_{ij} - \lambda_{0j}}{\alpha_j}\right|\right); \qquad with\ \alpha_j = \begin{cases}\lambda_{0j}\ \ if \quad \lambda_{0j} \neq 0 \\ 1\ \ otherwise\end{cases} \tag{41}$$

$$Std(\lambda_j) = \sqrt{\frac{\sum_{i=1}^{N}\left(\lambda_{ij} - \lambda_{Nj}\right)^2}{N}} \qquad with \quad \lambda_{Nj} = \frac{\sum_{i=1}^{N}\lambda_{ij}}{N} \tag{42}$$

The former quantifies the average magnitude of the estimator's accuracy for each parameter and the latter reflects the estimator's variation from sample to sample. Most of the investigations involving properties of multiparameter estimators have employed overall criteria. One commonly used measure of the goodness of a vector estimator is the relative mean square error

$$MSE(\lambda) = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{j=1}^{5}\left(\frac{\lambda_{ij} - \lambda_{0j}}{\alpha_j}\right)^2\right] \tag{43}$$

The statistics in (41)- (43) were calculated by generating N=1000 different random samples for each number of class, for each sample size, for each GLD model, and each estimation scheme. The number of replications of 1000 may arguably not be large enough to establish the actual values of Mrb, Std and MSE, but would suffice to assess shortcomings and relative merits of the various estimators.

The numerical optimization routines involved in the various methods of estimation require that initial guesses on the parameter estimates be made. In order to initialize the iteration process from a sufficiently good value we took a slight modification of the true values of the parameters

$$\lambda_j^0 = 0.001u + \lambda_{0\,j}[1 + 0.25u\,\mathrm{sgn}(v - 0.5)], \qquad j = 1, 2, \ldots, 5 \tag{44}$$

for each of the selected models. The quantities u and v are two independent random variables on (0,1).

Regrettably, it is possible for both the downhill simplex procedure and the Newton-Raphson method to approach a solution very far from λ_0. To reduce the impact of such behavior on performance comparisons, the optimization algorithms were initialized from five different starting points and the one corresponding to the lowest Euclidean distance from the true value was retained as the optimal solution.

The fast multiple recursive generator (FMRG) proposed by Deng and Lin (2000) was applied to produce uniform pseudorandom number on the unit interval. The choice of the starting values was based on the Wichmann-Hill algorithm (Wichmann and Hill, 1982). The seeds of the pseudo random sequences ensure that all the estimation schemes operate on exactly the same samples and start from the same initial parameter combinations. All the software has been written in Future Basic 7 running on a G4 (one processor, 1GHz) computer using Mac 10.3 operating system. Program codes as well as numerical results are available from the author on request.

Results

Table 2, in two parts, shows the relative mean bias (41) and the standard deviation (42) for the 36 configurations considered in the simulation plan. Actually, the two indices have been averaged over the five parameters to obtain summary measures. Table 3 reports the MSE obtained by the various methods.

Several points are worthy of note based on a thorough inspection of the findings in the tables.

1. The most obvious point to be noticed is that the GLD parameters were generally estimated with reasonable accuracy regardless the shape of the histogram of the generated data. Both the relative bias and the standard deviation decrease as the size of the samples increases (an indication that all of the methods provide consistent estimators). However, the general quality of the results rapidly deteriorates with increased population skewness. For the PLS method, all indicators achieve the best values. The MCM and PWM methods, in this order, have the worst performance among all six methods.

2. The relative performance of the methods did not greatly depend on the number of classes. This is surprising since a decrease in efficiency is expected in using fewer classes, at least for skewed distributions. The scarce or null influence of k on the behavior of the parameter estimators should be primarily ascribed to the large size of the samples used in our simulation plan. Also, the unbalanced partitions used for G(2,1) and B(2,4) have limited the negative effect of an increase of the variance due to an increase of k.

3. The method of percentiles has several advantages. For example, the outliers are given less weight than in the moment estimates; in fact, the PM estimators can still be computed when moments do not exist. Yet, the validity of this method is seriously hampered by the lack of a theoretical justification in selecting a given set of percentiles. The sextiles employed in our experiments have given satisfactory results for symmetric distributions, but the performance dropped down for the positively skewed distributions. Gaswirth,1972 pointed out that the practice of equally spaced percentiles is not an optimal choice for highly skewed distributions.

4. The method of moments produced good estimates in our experiments, but it is not recommendable for a general use with grouped data from continuous distributions. In part because, to operate at its best, the method requires corrections which are difficult to establish and in part because of the numerical and statistical instability of high order moments. Also, there is arbitrariness in the choice of the moments to equate.

5. The probability-weighted moment matching is not very accurate in estimating the GLD parameters for grouped data. However, since the PWM estimates can be calculated with relative easy, they can serve as a good automated routine to determine initial guesses for the parameters to be used with other methods.

6. The Cramér-von Mises method yields reliable estimates (at least for symmetrical distributions), but is hard to apply because of the intricate form of the Hessian matrix involved in the iteration process. In fact, the MCM method encountered the worst convergence problems of any of the techniques considered.

7. The scoring method employed to determine the maximum likelihood estimates has given results which are slightly better than those obtained by MCM, but not by an amount of likely practical importance. Both ML a d MCM are computationally demanding because of the complex and unwieldy objective function, but the former offers a well established theoretical background. In fact, it is common to use the asymptotic properties of the ML estimates to construct confidence regions for parameters.

8. In the vast majority of cases, the pseudo least squares method provided the best result. In this sense, we suggest PLS the as the candidate procedure to estimate the parameters of a GLD from grouped data. An attractive property of this method is the possibility that the initial guesses must be supplied only for the exponential parameters. Furthermore, the PLS method has the practical advantage of being simply executable using a nonlinear regression routine implemented in several statistical packages. On the other hand, the sample and asymptotic properties of the PLS method have as yet to be extensively explored. More theoretical work needs to be done.

Table 2a: mean relative bias for the 36 experiments

Model	k	n	PM	MM	PWM	MCM	ML	PLS
N(0,1)	8	1000	0.0715	0.1136	0.1256	0.1273	0.1660	0.1170
		2000	0.0590	0.0985	0.1191	0.1138	0.1088	0.0987
		4000	0.0509	0.0875	0.1130	0.1105	0.0994	0.0857
	14	1000	0.0709	0.1131	0.1243	0.1236	0.1479	0.1153
		2000	0.0572	0.0980	0.1162	0.1145	0.1071	0.0990
		4000	0.0471	0.0878	0.1120	0.1168	0.1042	0.0839
	20	1000	0.0728	0.0955	0.1236	0.1229	0.1311	0.1083
		2000	0.0603	0.0955	0.1138	0.1146	0.1093	0.1202
		4000	0.0513	0.0879	0.1094	0.1186	0.0996	0.0866
B(2,2)	8	1000	0.1719	0.0912	0.0726	0.0928	0.0776	0.1100
		2000	0.1396	0.0709	0.0573	0.0866	0.0761	0.0958
		4000	0.1037	0.0547	0.0449	0.0829	0.0772	0.0840
	14	1000	0.1720	0.0890	0.0722	0.0914	0.0701	0.1013
		2000	0.1301	0.0688	0.0570	0.0834	0.0752	0.0907
		4000	0.0974	0.0482	0.0435	0.0816	0.0760	0.0802
	20	1000	0.1770	0.0905	0.0694	0.0910	0.0692	0.0995
		2000	0.1411	0.0672	0.0540	0.0869	0.0741	0.0924
		4000	0.1093	0.0487	0.0436	0.0845	0.0753	0.0895
G(2,1)	8	1000	0.1718	0.1010	0.1078	0.3146	0.2604	0.1034
		2000	0.1515	0.1012	0.0962	0.3091	0.2191	0.0874
		4000	0.1337	0.1027	0.0888	0.2942	0.1897	0.0763
	14	1000	0.1815	0.0993	0.1060	0.3016	0.2355	0.0936
		2000	0.1443	0.1019	0.0911	0.2885	0.1812	0.0803
		4000	0.1152	0.1013	0.0794	0.2819	0.1447	0.0709
	20	1000	0.1600	0.0999	0.1022	0.3049	0.3197	0.0900
		2000	0.1353	0.1007	0.0873	0.2826	0.2275	0.0782
		4000	0.1084	0.1009	0.0812	0.3150	0.1912	0.0702
B(2,4)	8	1000	0.1418	0.0946	0.0828	0.1173	0.0879	0.1044
		2000	0.1225	0.0960	0.0679	0.1122	0.0722	0.0937
		4000	0.1120	0.0954	0.0592	0.1045	0.0681	0.0854
	14	1000	0.1550	0.0953	0.0915	0.1128	0.0855	0.0958
		2000	0.1210	0.0955	0.0867	0.1054	0.0703	0.0875
		4000	0.0982	0.0942	0.0806	0.0985	0.0668	0.0812
	20	1000	0.1558	0.0949	0.0806	0.1103	0.0904	0.0951
		2000	0.1254	0.0957	0.0685	0.1035	0.0752	0.0876
		4000	0.1024	0.0962	0.0602	0.1020	0.0720	0.0839
Total			**4.2189**	**3.2732**	**3.0895**	**5.5026**	**4.4013**	**3.3228**

Table 2b: mean standard deviation for the 36 experiments

Model	k	n	PM	MM	PWM	MCM	ML	PLS
N(0,1)	8	1000	0.0959	0.2031	0.2253	0.202	0.283	0.2148
		2000	0.0867	0.1398	0.2190	0.186	0.183	0.1751
		4000	0.0765	0.0954	0.2093	0.190	0.166	0.1507
	14	1000	0.1160	0.1975	0.2238	0.199	0.276	0.2023
		2000	0.1031	0.1402	0.2142	0.194	0.186	0.1755
		4000	0.0940	0.1004	0.2090	0.193	0.175	0.1469
	20	1000	0.1222	0.1934	0.2244	0.198	0.270	0.1940
		2000	0.1149	0.1428	0.2085	0.189	0.200	0.2174
		4000	0.1114	0.1023	0.2044	0.193	0.175	0.1559
B(2,2)	8	1000	0.1096	0.0504	0.0439	0.055	0.045	0.0552
		2000	0.0862	0.0397	0.0342	0.052	0.044	0.0431
		4000	0.0644	0.0295	0.0267	0.050	0.045	0.0328
	14	1000	0.1077	0.0510	0.0433	0.055	0.041	0.0478
		2000	0.0806	0.0408	0.0341	0.050	0.044	0.0360
		4000	0.0616	0.0290	0.0260	0.049	0.044	0.0286
	20	1000	0.1105	0.0519	0.0420	0.054	0.041	0.0444
		2000	0.0880	0.0398	0.0324	0.052	0.043	0.0387
		4000	0.0667	0.0295	0.0260	0.051	0.044	0.0349
G(2,1)	8	1000	0.4391	0.8184	0.5738	1.014	0.957	0.5548
		2000	0.4305	0.8399	0.5572	1.035	0.879	0.4932
		4000	0.4024	0.8321	0.5517	1.030	0.830	0.5082
	14	1000	0.4730	0.8236	0.5654	1.167	0.918	0.5409
		2000	0.4352	0.8193	0.5516	1.134	0.793	0.5333
		4000	0.3904	0.8032	0.5364	1.103	0.740	0.4999
	20	1000	0.4538	0.8115	0.5583	1.067	1.011	0.5330
		2000	0.4188	0.8164	0.5444	1.068	0.879	0.5082
		4000	0.4012	0.8316	0.5545	1.101	0.813	0.4835
B(2,4)	8	1000	0.0656	0.0508	0.0432	0.065	0.043	0.0423
		2000	0.0532	0.0516	0.0355	0.073	0.037	0.0357
		4000	0.0437	0.0509	0.0308	0.059	0.034	0.0308
	14	1000	0.0831	0.0514	0.0430	0.062	0.042	0.0386
		2000	0.0668	0.0511	0.0383	0.059	0.035	0.0340
		4000	0.0539	0.0503	0.0345	0.056	0.033	0.0309
	20	1000	0.0848	0.0508	0.0423	0.061	0.045	0.0379
		2000	0.0694	0.0509	0.0353	0.057	0.038	0.0332
		4000	0.0578	0.0513	0.0308	0.056	0.036	0.0313
Total			**6.1187**	**9.5315**	**7.5736**	**12.480**	**10.468**	**6.9638**

Table 3: mean square error for the 36 experiments

Model	k	n	PM	MM	PWM	MCM	ML	PLS
N(0,1)	8	1000	0.0622	0.1343	0.1865	0.2437	0.3458	0.1591
		2000	0.0416	0.0925	0.1776	0.1792	0.1293	0.1160
		4000	0.0298	0.0664	0.1565	0.1548	0.1039	0.0854
	14	1000	0.0573	0.1229	0.1802	0.2099	0.2572	0.1569
		2000	0.0343	0.0892	0.1606	0.1810	0.1180	0.1164
		4000	0.0220	0.0665	0.1550	0.2315	0.1098	0.0806
	20	1000	0.0570	0.0870	0.1851	0.2013	0.2072	0.1438
		2000	0.0375	0.0844	0.1552	0.1710	0.1272	0.2057
		4000	0.0275	0.0652	0.1423	0.2016	0.1006	0.0897
B(2,2)	8	1000	0.3041	0.0897	0.0442	0.0716	0.0473	0.0967
		2000	0.1930	0.0582	0.0277	0.0641	0.0443	0.0737
		4000	0.1043	0.0350	0.0173	0.0586	0.0451	0.0576
	14	1000	0.2854	0.0868	0.0432	0.0700	0.0395	0.0802
		2000	0.1634	0.0561	0.0272	0.0594	0.0437	0.0644
		4000	0.0928	0.0289	0.0162	0.0578	0.0437	0.0519
	20	1000	0.3097	0.0895	0.0404	0.0681	0.0387	0.0741
		2000	0.1985	0.0533	0.0242	0.0633	0.0428	0.0652
		4000	0.1114	0.0294	0.0160	0.0604	0.0433	0.0616
G(2,1)	8	1000	0.6005	0.0752	0.1291	2.6568	2.3653	0.1258
		2000	0.4633	0.0753	0.0989	2.9661	1.6327	0.0815
		4000	0.3214	0.0768	0.0822	2.4534	0.9792	0.0560
	14	1000	0.6473	0.0728	0.1249	2.6114	2.0890	0.0936
		2000	0.4243	0.0764	0.0920	2.5488	0.9789	0.0619
		4000	0.2499	0.0755	0.0691	2.1998	0.4114	0.0476
	20	1000	0.5364	0.0739	0.1165	2.0654	3.7986	0.0853
		2000	0.3625	0.0745	0.0822	1.6519	1.5767	0.0591
		4000	0.2147	0.0751	0.0686	2.4292	0.9067	0.0460
B(2,4)	8	1000	0.2052	0.0659	0.0618	0.1287	0.0673	0.0848
		2000	0.1439	0.0676	0.0414	0.1288	0.0457	0.0694
		4000	0.1133	0.0669	0.0309	0.1031	0.0393	0.0601
	14	1000	0.2639	0.0672	0.0732	0.1158	0.0627	0.0727
		2000	0.1608	0.0672	0.0659	0.1019	0.0431	0.0620
		4000	0.0961	0.0655	0.0555	0.0905	0.0390	0.0545
	20	1000	0.2667	0.0668	0.0581	0.1079	0.0693	0.0711
		2000	0.1711	0.0669	0.0418	0.0980	0.0479	0.0622
		4000	0.1099	0.0681	0.0326	0.0945	0.0440	0.0579
Total			**7.4830**	**2.6132**	**3.0801**	**24.8991**	**17.0844**	**3.0305**

References

Berkson J. (1981): Minimum Chi-square, not maximum likelihood! *Annals of statistics*, 8, 457-487.

Bonferroni C. E. (1940-41): Elementi di statistica generale. Litografia Gili, Torino.

Castillo E., Hadi A.S.(1995): A method for estimating parameters and quantiles of distributions of continuous random variables. *Computational Statistics and Data Analysis*, 20, 421-439.

Cheng R.C.H., Iles T.C. (1987): Corrected maximum likelihood in non-regular problems. *Journal of the Royal Statistical Society*, B, 49, 95-101.

David H. A., Mishriky R. S. (1968): Order statistics for discrete population and for grouped samples. *Journal of the American Statistical Association*, 63, 1390-1398.

Deng L. Y., Lin D.K.J. (2000): Random number generation for the new century. *The American Statistician*, 54, 45-150.

Devroye L. (1996). Random variate generation in one line of code. *Proceedings of the Winter Simulation Conference*, 265-272.

Duchesne T, Rioux J., Luong A. (1997): Minimum Cramér-Von Mises distance methods for complete and grouped data. *Communication in Statistics: Theory and Methods*, 26, 401-420.

Eubank R.L. (1988): Optimal grouping, spacing, stratification, and piecewise constant approximation, *SIAM Review*, 30, 404-420.

Faure H. (1982): Discrépance de suites associées à un système de numération (en dimension s*). Acta Aritmetica*, 41, 337-351.

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

Fox B. L. (1986): Implementation and relative efficiency of quasi-random sequence generators. *ACM Transactions on Mathematical Software*, 12, 362-376

Freimer M., Mudholkar S., Kollia G., Lin T.C. (1988): A study of the generalized Tukey Lambda family. *Communication in Statistics: Theory and Methods*, 17, 3547-3567.

Gastwirth J.L. (1972): The estimation of Lorenz curve and Gini index, *Review of Economics and Statistics*, 52, 306-316.

Gilchrist W. (2000): Statistical modeling with quantile function. Chapman & Hall, CRC, Boca Raton (FL).

Greenwood J. A., Landwehr J. M., Matals N. C., Wallis J. R. (1979): Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resource Research*, 15, 1049-1054.

Groeneveld R. A. (1986): Measurement of skewness and kurtosis for the generalized Tukey lambda distributions. *Communication in Statistics - Theory and Methods*, 15, 329-343.

Harrell F.E. Davis C.E. (1982): A new distribution free quantile estimator, *Biometrika*, 69, 635-640.

Hoaglin D.C. (1975): The small-sample variance of the Pitman location estimators. *Journal of the American Statistical Association*, 70, 880-888.

Joiner B.L., Rosenblatt J. R. (1971): Some properties of the range in samples from Tukey's Symmetric lambda distribution. *Journal of the American Statistical Association*, 66, 394-399.

Karian Z. E., Dudewicz E. J., McDonald P. (1996): The extended generalized lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the 'final' word on moment fits. *Communication in Statistics: Simulation and Computation*, 25, 611-642.

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

Karian Z.A., Dudewicz E. J. (2000): Fitting statistical distributions. The generalized lambda distribution and generalized bootstrap methods. CRC, Boca Raton (FL).

Kariya T. (1986): Analogous t and f test statistics based on grouped data. *Pacific Statistical Congress*, I.S. Francis, B.F.J. Manly, and F.C. Lam (Eds.). Elsevier Science Publishers B.V. (North-Holland)

King A.R., MacGillivray H. L. (1999): A starship estimation for the generalized lambda distribution. *Australian and New Zealand Journal of Statistics*, 41, 353-374.

Korn E.L., Miothorne D., Graubard B.J. (1997): Estimating interpolated percentiles from grouped data with large samples, *Journal of Official Statistics*, 13, 385-399

Lakhany A., Mausser H. (2000): Estimating the parameters of the generalized Lambda distribution. *Algo Research Quarterly*, 3, 47-58.

Lam H., Bowman K. O., Shenton L .R. (1980): Remarks on the generalized Tukey's lambda family of distributions. *Proceedings of the ASA, Statistical Computing Section*, Houston, Texas, August 11-14, 1980, 134-139.

LaRiccia V. (1982): Asymptotic properties of weighted L^2 quantile distance estimators. *The Annals of Statistics*. 10, 621-624.

Lawton W. H., Sylvestre E. A. (1971): Elimination of linear parameters in nonlinear regression. *Technometrics*, 13, 461-467.

MacGillivray H.L. (1982). Skewness properties of asymmetric forms of Tukey lambda distributions. *Communication in Statistics: Theory and methods*, 11, 2239-2248.

MacDonald J.B., Ransom M.R. (1979): Alternative parameter estimators based upon

grouped data. Communication *in Statistics - Theory and Methods*, A8, 899-917.

Nelder J. A., Mead R. (1965): A simplex method for function minimization. *The Computer Journal*, 7, 308-313.

Olsson D. M., Nelson L. S. (1975): The Nelder-Mead simplex procedure for function minimization. *Technometrics*, 17, 45-51.

Oztürk A., Dale R.F. (1985): Least squares estimation of the parameters of the generalized lambda distributions. *Technometrics*, 27, 81-84.

Oztürk A., Abouammouh A. M. (1987): On fitting the generalized lambda distribution to Climatological data. *Pakistan Journal of Statistics, B*, 3, 39-50.

O'Neill B., Wells W.T. (1972): Some recent results in lognormal parameter estimation using grouped and ungrouped data. *Journal of the American Statistical Association*, 67, 76-80.

Parr W. C. (1981). Minimum distance estimation: survey and bibliography. *Communications in Statistics, A*, 10, 1205-1224.

Press W.H., Teukolsky S.A.,Vetterling W.T.,Flannery B.P. (1993): Numerical recipes in C: The art of scientific computing. Cambridge University Press, Cambridge.

Ramberg J.S., Schmeiser B.W. (1974): An approximate method for generating asymmetric random variables. *Communication of the ACM*, 17, 78-82.

Ramberg J. S., Dudewicz E. J., Tadikamalla P. R., Mykytka E. F. (1979): A probability distribution and its uses in fitting data. *Technometrics*, 21, 201-214.

Ramberg J. S., Dudewicz E. J., Tadikamalla P. R., Mykytka E. F. (1980): Letters to the Editor. *Technometrics*, 22, 135.

Rao C.R. (1973): Linear statistical inference and its applications (2nd ed.). John Wiley & Sons, New York.

Samuel E., Bachi R. (1964): Measure of distance of distribution function and some applications. *Metron*, 23, 83-112.

Sarabia J. M. (1996): A hierarchy of Lorenz curves based on generalized Tukey's lambda distribution. *Econometric Reviews*, 16, 305-320.

Shapiro S. S., Wilk M. B. (1965): An analysis of variance test for normality (complete samples). *Biometrika*, 52,591-611.

Schmeiser, B. W., Deutsch, S. J. (1977): Quantile estimation from grouped data: The cell midpoint. *Communications in Statistics: Simulation and Computation, B6*, 221-234.

Wichman B.A., Hill. I.D. (1982): An efficient and portable pseudo-random number generator. *Applied Statistics*. Vol. 31,188-190.