Statistical Papers 43, 237–255 (2002)

Statistical Papers

Parameter estimation with grouped data according to the linearization method – a comparison with alternative approaches

Max D. Jöhnk and Stefan Niermann

Institute for Quantitative Economics, University of Hannover,
Koenigsworther Platz 1, 30167 Hannover, Germany

Received: January 11, 2000; revised version: August 23, 2000

Key Words: estimation; grouped data; linearization.

This paper considers the problem of parameter estimation when data of a random sample are given in the form of a frequency table. We give special consideration to a method that linearizes the cumulative distribution function (CDF). In that case parameters can be derived from the weighted estimation of a linear regression equation. The favourable properties of this estimation technique are demonstrated in a simulation experiment, where the parameters of a two-parameter-Weibull distribution are estimated.

1. INTRODUCTION

In the process of working with data one often faces the problem that the observations are only available in a grouped form. If one is interested in estimating the unknown paramaters of a distribution it is possible to apply the Maximum-Likelihood-Estimator for grouped data or the Minimum-Chi-Square-Estimator. These two methods are described by McDONALD and RANSOM (1979). Another possible method is the quantile estimator that can be extended to the Minimum-Quantile-Distance-Estimator (JOHNSON, KOTZ and BALAKRISHNAN (1994)). An excellent additional reference to these estimation procedures is given by COX and HINKLEY (1974, p. 306).

We present a new estimation method which can be applied to the most often assumed two-parameter distribution models. This method is numerically easy to perform, yields a closed form solution and has good properties. In section 2 the several estimation methods are described. In the third section these methods are compared with regard to their asymptotic properties. In order to be able to derive properties when the sample size is finite, simulations are performed in section 4.

2. ESTIMATION METHODS

With given classes $\tilde{x}_{i-1} < X \leq \tilde{x}_i$ for $i = 1, \dots k$ the counts n_i in a frequency table are random variables. The relative frequencies are denoted with $\hat{p}_i = n_i/(\sum_{i=1}^k n_i) = n_i/n$ and the cumulated relative frequencies are denoted with $\hat{F}_i$ for $i = 1, \dots, k$ with $\hat{F}_k = 1$.

The distribution of X depends upon an unknown parameter vector, which is denoted with $\boldsymbol{\theta}$. The probabilities $p_i := P(\tilde{x}_{i-1} < X \leq \tilde{x}_i)$ and $F_i = P(X \leq \tilde{x}_i)$ are functions of $\boldsymbol{\theta}$ as well. The supplement $(\boldsymbol{\theta}\ )$ is omitted unless this fact is to be stressed. Vectors and matrices are set in bold letters.

The $\tilde{x}_i$ can be interpreted as empirical quantiles. However, note that in a situation with given class breaks not the quantiles $\hat{Q}(w)$, but the corresponding w depend upon the random sample.

The alternative estimation methods are analyzed and compared:

a) Maximum-Likelihood-Method
Approach:

Choose the parameter vector $\boldsymbol{\theta}$ such that

$$L(\boldsymbol{\theta}\ ) = \left(\frac{n!}{\prod_i (n_i!)}\right) \cdot \prod p_i(\boldsymbol{\theta}\ )^{n_i} \tag{1}$$

is maximized.

b) Minimum-Chi-Square-Method
Approach:

Choose the parameter vector $\boldsymbol{\theta}$ such that

$$X^2 = \sum \frac{[n_i - np_i(\boldsymbol{\theta}\ )]^2}{np_i(\boldsymbol{\theta}\ )} \tag{2}$$

is minimized.

c) Optimal fit of cumulative distribution function
Approach:

Choose the parameter vector $\boldsymbol{\theta}$ such that

$$S_1^2 = \sum_i \sum_j [\hat{F}_i - F_i(\boldsymbol{\theta}\ )][\hat{F}_j - F_j(\boldsymbol{\theta}\ )]w_{ij} = [\hat{\mathbf{F}} - \mathbf{F}(\boldsymbol{\theta}\ )]'\mathbf{W}_1[\hat{\mathbf{F}} - \mathbf{F}(\boldsymbol{\theta}\ )] \tag{3}$$

is minimized. The inverse of the variance matrix
$\mathbf{V}_{\hat{\mathbf{F}}} = E[(\hat{\mathbf{F}} - \mathbf{F}(\boldsymbol{\theta}))(\hat{\mathbf{F}} - \mathbf{F}(\boldsymbol{\theta}))']$ is an appropriate choice for $\mathbf{W}_1$. This matrix is given in section 3.2.

d) Optimal fit of cumulative distribution function after linearization

Method c) corresponds to the adaption of a nonlinear regression function to the empirical CDF. Often it is possible to linearize this regression function and consequently to adapt calculation methods for the parameter estimators in the linear regression model.
The linearization is carried out using appropriate transformation functions $g(\cdot)$ and $h(\cdot)$, which enable us to write the CDF in the form

$$h(F(x)) = \beta_0 + \beta_1 \cdot g(x).$$

The idea is to plot $g(\tilde{x}_i)$ against $h(\hat{F}(\tilde{x}_i))$. These points should scatter around a straight line if the assumed distribution is an appropriate model for the data. The estimates of the regression parameters can be utilized for estimating the unknown model parameters. Additionally to parameter estimation, this approach can be used as a check for model validity.

Hence, the parameters β_0 and β_1 are chosen such that

$$\begin{aligned} S_2^2 &= \sum_i \sum_j [h(\hat{F}_i) - \beta_0 - \beta_1 \cdot g(x_i)] w_{ij}^2 [h(\hat{F}_j) - \beta_0 - \beta_1 \cdot g(x_j)] \quad (4) \\ &= [h(\hat{\mathbf{F}}) - h(\mathbf{F}(\beta_0, \beta_1))]' \mathbf{W}_2 [h(\hat{\mathbf{F}}) - h(\mathbf{F}(\beta_0, \beta_1))] \end{aligned}$$

is minimized. Here the function h is applied elementwise to the vectors $\mathbf{F} = \mathbf{F}(\boldsymbol{\theta})$ and $\hat{\mathbf{F}} = \hat{\mathbf{F}}(\boldsymbol{\theta})$.

The inverse of the variance matrix

$$\mathbf{V}_h = E[h(\hat{\mathbf{F}}) - h(\mathbf{F}(\beta_0, \beta_1))][h(\hat{\mathbf{F}}) - h(\mathbf{F}(\beta_0, \beta_1))]']$$

is an appropriate choice for $\mathbf{W_2}$, which is given in section 3.3.

Some examples for the linearization of the CDF follow.

(i) The normal distribution
For the normal distribution we have

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

and

$$\Phi^{-1}(F(x)) = -\frac{\mu}{\sigma} + \frac{1}{\sigma} x$$

In this case $h(y)$ is the inverse function of the CDF of the standard normal distribution and $g(x) = x$. Note that the parameters μ and σ of the normal distribution do not coincide with the regression coefficients. We rather have $\beta_0 = -\mu/\sigma$ and $\beta_1 = 1/\sigma$. It follows $\hat{\sigma} = 1/\hat{\beta}_1$ and $\hat{\mu} = -\hat{\beta}_0/\hat{\beta}_1$.

The normal probability paper as described by JOHNSON, KOTZ and BALAKRISHNAN (1994) is based on this transformation.

(ii) Just like in the case of the normal distribution there are many other distributions with $g(x) = x$. Distributions of this type are said to have a linear parametric distribution. Examples are:

(iia) The two-parameter exponential distribution:

$$F(x) = 1 - exp\{-\lambda(x - \xi)\}$$
$$\ln(1 - F(x)) = \lambda\xi - \lambda x$$

It follows $h(y) = \ln(1 - y)$.

(iib) The logistic distribution:

$$F(x) = \frac{exp\left(\frac{x-m}{d}\right)}{1 + exp\left(\frac{x-m}{d}\right)}$$
$$\ln\left(\frac{F(x)}{1 - F(x)}\right) = -\frac{m}{d} + \frac{1}{d}x$$

It follows $h(y) = \ln\left(\frac{y}{1-y}\right)$.

(iic) The Gumbel distribution (Extreme value distribution):

$$F(x) = 1 - exp\left(-exp\left(\frac{x - m}{d}\right)\right)$$
$$\ln(-\ln(1 - F(x))) = -\frac{m}{d} + \frac{1}{d}x$$

It follows $h(y) = \ln(-\ln(1 - y))$.

(iid) The Cauchy distribution:

$$F(x) = \frac{1}{\pi} arc\,tan\left(\frac{x - m}{d}\right) + \frac{1}{2}$$
$$tan(\pi(F(x) - \frac{1}{2})) = -\frac{m}{d} + \frac{1}{d}x$$

where $h(y) = tan(\pi(y - \frac{1}{2}))$.

(iie) The Perk distribution (Burr-VIII distribution with $k = 1$):

$$F(x) = \frac{2}{\pi} \cdot arctan\left(exp\left(\frac{x - \xi}{\lambda}\right)\right)$$
$$\ln(\tan(\frac{\pi}{2} \cdot F(x))) = -\frac{\xi}{\lambda} + \frac{1}{\lambda} \cdot x$$

It follows $h(y) = \ln(tan(\pi y/2))$.

(iii) Many distributions have the property that the logarithm of the considered random variable has a linear parametric distribution: $g(x) = \ln(x)$. A well-known example for distributions of this type is the Gibrat distribution.

(iiia) The Gibrat distribution (Lognormal distribution):

$$F(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right)$$

$$\Phi^{-1}(F(x)) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}\ln x$$

(iiib) The Pareto distribution (Logexponential distribution):

$$F(x) = 1 - \left(\frac{b}{x}\right)^p$$

$$\ln(1 - F(x)) = p \cdot \ln b - p \cdot \ln x$$

This type of linearization is utilized in the Pareto diagram.

(iiic) The Fisk distribution (Loglogistic distribution):

$$F(x) = \frac{x^p}{c^p + x^p}$$

$$\ln\left(\frac{F(x)}{1 - F(x)}\right) = -p \cdot \ln c + p \cdot \ln x$$

(iiid) The two-parameter Weibull distribution (Loggumbel distribution):

$$F(x) = 1 - exp(-ax^b)$$

$$\ln(-\ln(1 - F(x))) = \ln a + b \cdot \ln x$$

The functions $h(\cdot)$ of the cases iiia) - iiid) coincide with those of the cases i) and iia) - iic).

(iv) Distributions with three parameters can often be linearized with respect to the unknown parameters if one parameter is known. Some examples follow.

(iva) The three-parameter Weibull distribution:

$$F(x) = 1 - exp(-a \cdot (x - c)^b)$$

If c is known, this case obviously coincides with case iiid). However, if b is known, the parameters a and c can be estimated from the linearization

$$(-\ln(1 - F(x)))^{1/b} = a^{1/b} \cdot (x - c) = -c \cdot a^{1/b} + a^{1/b} \cdot x$$

with $h(y) = (-\ln(1 - y))^{1/b}$.

(ivb) The case of the lognormal distribution with three parameters is similar.

$$F(x) = \Phi\left(\frac{\ln(x-\xi)-\mu}{\sigma}\right)$$

In analogy to case iva) the estimation problem reduces to case iiia) if ξ is known.

If σ is known, the parameters ξ and μ can be estimated from the linearization

$$exp(\sigma \cdot \Phi^{-1}(F(x))) = -\xi \cdot exp(-\mu) + exp(-\mu) \cdot x$$

with $h(y) = exp(\sigma \cdot \Phi^{-1}(y)$.

(ivc) Somewhat different is the case with the scale parametrized Burr-XII-distribution.

$$F(x) = 1 - \frac{1}{\left(1+\left(\frac{x}{\lambda}\right)^c\right)^k}$$

If k is known, c and λ can be estimated from the linearization

$$\ln\left((1-F(x))^{-1/k}\right) = -c \cdot \ln(\lambda) + c \cdot \ln(x)$$

with $h(y) = \ln((1-y)^{-1/k} - 1)$.

If c and λ are known, it is possible to estimate k from the linearization

$$-\ln(1-F(x)) = k \cdot \ln\left(1+\left(\frac{x}{\lambda}\right)^c\right)$$

in a linear regression without intercept.

(ivd) This also applies to the scale parametrized Burr-III-distribution.

$$F(x) = \left(1+\left(\frac{\lambda}{x}\right)^c\right)^{-k}$$

If k is known, c and λ can be estimated from the linearization

$$\ln(F(x)^{-1/k} - 1) = c \cdot \ln(\lambda) - c \cdot \ln(x)$$

with $h(y) = \ln(y^{-1/k} - 1)$.

If c and λ are known, it is possible to estimate k from the linearization

$$-\ln(F(x)) = k \cdot \ln\left(1+\left(\frac{\lambda}{x}\right)^c\right)$$

in a linear regression without intercept.

(v) Further possibilities of linearization can be derived for distributions as introduced by JOHNSON (1949) and for Burr-distributions of the types V and VI.

e) Minimum-Quantile-Distance-Method

Choose the parameter vector $\boldsymbol{\theta}$ such that

$$S_3^2 = [\hat{\mathbf{Q}} - \mathbf{Q}(\boldsymbol{\theta})]'\mathbf{W}_3[\hat{\mathbf{Q}} - \mathbf{Q}(\boldsymbol{\theta})] \tag{5}$$

is minimized. A matrix of weights was proposed by CARMODY, EUBANK and LARICCIA (1984).

Further methods of estimation are not considered because they cannot be classified. Notable are the approaches described by LAWLESS (1982) and CHENG and CHEN (1988). These approaches are based on the linearization of the survival respectively the hazard function of grouped lifetime data. Insofar these approaches are similar to method d). These approaches are based on the assumption that the hazard rate remains constant within the classes. This (usually violated) assumption is not needed when method d) is applied.

The unweighted versions of the methods c) to e) can be considered as simplifications of the weighted versions.

All the methods described above are based on the minimization or the maximization of an objective function. It is clear that two methods using the same objective function will render identical parameter estimates. In case the objective function of one method is a monotone function of the objective function of another method, both methods will render identical parameter estimates, too. If two objective function converge to the same function as n increases, the parameter estimates will coincide asymptotically (asymptotic equivalence).

3. A COMPARISON OF METHODS OF ESTIMATION

3.1 ML - ESTIMATE AND MINIMUM - χ^2 - ESTIMATE

The ML-estimate uses the objective function:

$$L = \left(\frac{n!}{\prod_i (n_i!)}\right) \cdot \prod p_i(\boldsymbol{\theta})^{n_i} \qquad \text{with} \quad p_i = p_i(\boldsymbol{\theta})$$

with the following equivalent (monotonely transformed) function

$$\ln \bar{L} = \sum n_i \cdot \ln p_i - \sum n_i \cdot \ln \hat{p}_i$$

Because the second term does not depend upon $\boldsymbol{\theta}$, $\bar{L}$ has a minimum at the same position as L. It follows:

$$\ln \bar{L} = \sum n_i \ln \frac{p_i n}{n_i}$$

For $\ln(x)$ we have the approximation $\ln(x) \approx (x^2 - 1)/2x$ with very good approximation results when $x \approx 1$. Therefore,

$$\ln \bar{L} \approx \sum n_i \left(\frac{p_i^2 n^2}{n_i^2} - 1 \right) \frac{n_i}{2np_i}$$
$$= \frac{n}{2} - \frac{1}{2n} \sum \frac{n_i^2}{p_i} = -\frac{1}{2} X^2$$

First, we have that the objective functions L and χ^2 are approximately equivalent, where the approximation quality improves as $\frac{n_i}{n}$ approaches $p_i(\boldsymbol{\theta})$. With a correct model specification one will be able to find parameter values $\hat{\boldsymbol{\theta}}$ such that $plim \frac{n_i}{n} = p_i(\hat{\boldsymbol{\theta}})$ and accordingly both methods will yield asymptotically identical estimators. With an incorrect specification we still have $plim \frac{n_i}{n} = p_i$, but it won´t be possible to represent $p_i = p_i(\boldsymbol{\theta})$.

3.2 MINIMUM - χ^2 ESTIMATOR AND OPTIMAL FIT OF THE CUMULATIVE DISTRIBUTION FUNCTION

It can be shown that both methods use the same objective function and therefore yield identical estimates.

The comparison is complicated by the fact that the minimum - χ^2 estimator compares k relative frequencies $\hat{p}_i$ to k probabilities $p_i(\boldsymbol{\theta})$, whereas only $k-1$ cumulated pairs $\hat{F}_i$, $F_i(\boldsymbol{\theta})$ are used for the optimal fit of the cumulated distribution function. Indeed both the vectors $\hat{\mathbf{p}} = (\hat{p}_1, \ldots, \hat{p}_k)$ and $\mathbf{p}(\boldsymbol{\theta}) = (p_1(\boldsymbol{\theta}), \ldots, p_k(\boldsymbol{\theta}))$ have one superfluous element that can be determined from the other $k - 1$ elements. It follows:

$$\hat{p}_k = 1 - \sum_{1}^{k-1} p_i$$

$$p_k(\boldsymbol{\theta}) = 1 - \sum_{i=1}^{k-1} p_i(\boldsymbol{\theta})$$

Nevertheless X^2 uses all the elements of the vectors $\hat{\mathbf{p}}$ and $\mathbf{p}(\boldsymbol{\theta})$.

$$\begin{aligned}
X^2 &= \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = n \sum_{i=1}^{k} \frac{(\hat{p}_i - p_i)^2}{p_i} \\
&= n \sum_{i=1}^{k-1} \frac{(\hat{p}_i - p_i)^2}{p_i} + n \frac{[(1 - \sum_{i=1}^{k-1} \hat{p}_i) - (1 - \sum_{i=1}^{k-1} p_i)]^2}{1 - \sum_{i=1}^{k-1} p_i} \\
&= n \left\{ \sum_{i=1}^{k-1} \frac{(\hat{p}_i - p_i)^2}{p_i} + \frac{\left(\sum_{i=1}^{k-1} p_i - \sum_{i=1}^{k-1} \hat{p}_i \right)^2}{1 - \sum_{i=1}^{k-1} p_i} \right\} \\
&= n \left\{ \sum_{i=1}^{k-1} \frac{(\hat{p}_i - p_i)^2}{p_i} + \frac{\left[\sum_{i=1}^{k-1} (\hat{p}_i - p_i) \right]^2}{1 - \sum_{i=1}^{k-1} p_i} \right\}
\end{aligned}$$

For the rest of this section the first $k-1$ elements of $\hat{\mathbf{p}}$ and $\mathbf{p}(\boldsymbol{\theta})$ will be denoted by $\hat{\mathbf{p}}$ and $\mathbf{p}(\boldsymbol{\theta})$.

Hence, in matrix representation we have:
Let $\mathbf{1}$ be a vector of ones. Consequently $\mathbf{11}'$ is a square matrix of ones. Let further be $\check{\mathbf{x}}$ a diagonal matrix set up from the elements of the vector $\mathbf{x}$. It follows:

$$\frac{1}{n}X^2 = (\hat{\mathbf{p}} - \mathbf{p})' \left[(\check{\mathbf{p}})^{-1} + \frac{\mathbf{11}'}{p_k} \right] (\hat{\mathbf{p}} - \mathbf{p})$$

Note that the vectors $\mathbf{p}$ and $\hat{\mathbf{p}}$ consist of $k-1$ elements and that $p_k = 1 - \sum_{i=1}^{k-1} p_i = 1 - \mathbf{1}'\mathbf{p}$. The matrix put in brackets [...] is by the way the inverse of the variance-matrix of $\hat{\mathbf{p}}$:

$$\mathbf{V}_{\hat{\mathbf{p}}} = E(\hat{\mathbf{p}} - \mathbf{p})(\hat{\mathbf{p}} - \mathbf{p})' = \check{\mathbf{p}} - \mathbf{pp}'.$$

This can easily be seen by

$$(\check{\mathbf{p}} - \mathbf{pp}')((\check{\mathbf{p}})^{-1} + \frac{\mathbf{11}'}{p_k})$$

$$= \mathbf{I} - \mathbf{p1}' + \frac{\mathbf{p1}'}{p_k} - \frac{\mathbf{p}(1-p_k)\mathbf{1}'}{p_k} = \mathbf{I} \quad q.e.d.$$

It follows the representation:

$$\frac{X^2}{n} = (\hat{\mathbf{p}} - \mathbf{p})' \mathbf{V}_{\hat{\mathbf{p}}}^{-1} (\hat{\mathbf{p}} - \mathbf{p})$$

From this representation the variance matrix $V_{\hat{F}}$ and its inverse can be derived.

Because

$$\mathbf{K} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

(cumulation matrix)

$$(\hat{\mathbf{F}} - \mathbf{F}) = \mathbf{K}(\hat{\mathbf{p}} - \mathbf{p})$$

and

$$\mathbf{V}_{\hat{\mathbf{F}}} = E[(\hat{\mathbf{F}} - \mathbf{F})(\hat{\mathbf{F}} - \mathbf{F})'] = E[\mathbf{K}(\hat{\mathbf{p}} - \mathbf{p})(\hat{\mathbf{p}} - \mathbf{p})'\mathbf{K}'] = \mathbf{K}\mathbf{V}_{\hat{p}}\mathbf{K}'$$

it follows

$$\mathbf{V}_{\hat{\mathbf{F}}}^{-1} = \mathbf{K}'^{-1}\mathbf{V}_{\hat{p}}^{-1}\mathbf{K}^{-1},$$

where

$$\mathbf{K}^{-1} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

(decumulation matrix). Therefore,

$$\frac{X^2}{n} = (\hat{\mathbf{p}} - \mathbf{p})'\mathbf{V}_{\hat{p}}^{-1}(\hat{\mathbf{p}} - \mathbf{p}) = (\hat{\mathbf{p}} - \mathbf{p})'\mathbf{K}'\mathbf{K}'^{-1}\mathbf{V}_{\hat{p}}^{-1}\mathbf{K}^{-1}\mathbf{K}(\hat{\mathbf{p}} - \mathbf{p})$$

$$= (\hat{\mathbf{F}} - \mathbf{F})'\mathbf{V}_{\hat{F}}^{-1}(\hat{\mathbf{F}} - \mathbf{F})'$$

The estimation methods are identical if the approriate weighting matrix is used. Note, that $\mathbf{V}_{\hat{F}}^{-1}$ is the appropriate weight matrix $\mathbf{W}_1$ in equation 3.

3.3 OPTIMAL FIT OF THE CUMULATIVE DISTRIBUTION FUNCTION WITH AND WITHOUT LINEARIZATION

An approximation of the the linear regression equation $v_i = h(\hat{F}_i) - h(x_i; \beta_0, \beta_1)$ can be obtained by developping the disturbance terms of the nonlinear regression equation

$$u_i = \hat{F}_i - F(x_i; \beta_0, \beta_1)$$

in a Taylor expansion:

$$h(F(x_i; \beta_0, \beta_1)) = h(\hat{F}_i - u_i).$$

This yields

$$\begin{aligned} h(\hat{F}_i) &= h(F(x_i;\beta_0,\beta_1)) + (\hat{F}_i - F(x_i;\beta_0,\beta_1)) \cdot h'(F(x_i;\beta_0,\beta_1)) + \ldots \\ &= h(F(x_i;\beta_0,\beta_1)) + u_i \cdot h'(F(x_i;\beta_0,\beta_1)) + \ldots \\ &= h(F(x_i;\beta_0,\beta_1)) + u_i \cdot h'(\hat{F}_i) + \ldots \\ &= a + b \cdot g(x_i) + u_i \cdot h'(\hat{F}_i) + \ldots \end{aligned}$$

With $n \to \infty$ the approximation error converges to *zero*.

Let $\mathbf{V}_U$ denote the variance matrix of the disturbance vector $\mathbf{u}$. It coincides with the variance matrix $\mathbf{V}_{\hat{\mathbf{F}}}$. Then the variance matrix $\mathbf{V}_V$ of the disturbances $\mathbf{v}$ is approximately given by

$$\mathbf{V}_V = diag(h'(\hat{F}_i))\mathbf{V}_U diag(h'(\hat{F}_i)) \qquad (6)$$

This matrix coincides with the variance matrix $\mathbf{V}_h$ and therefore is an appropriate matrix of weights $\mathbf{W}_2$ in equation 4.

The objective function S_2^2 differs only insofar from the objective function S_1^2 as the disturbances cannot be transferred exactly in the linearization procedure. With a growing number of observations this effect of the linearization converges to *zero*. The methods described in c) and e) are therefore asymptotically equivalent.

3.4 OPTIMAL FIT OF THE CUMULATIVE DISTRIBUTION FUNCTION AND MINIMUM - QUANTILE - DISTANCE - METHOD

In case the quantile function is linear in the coefficients - and that is always the case if the cumulative distribution function can be linearized - the minimum-quantile-distance-method is just the estimation of the inverse regression equation compared to the linearized cumulative distribution function. This inversion leads to a modification of the regression results. But given a correct specification of the model, the observations are located on a straight line as n converges to infinity. Therefore, the estimation results have to coincide asymtotically.

In case the quantile function can not be linearized, this argumentation still applies. As above, the minimization of the quantile distances is the inversion of the optimal fit of the cumulative distribution function. Therefore, the estimation differences have to disappear as n converges to infinity if the model is specified correctly.

4. AN APPLICATION: THE TWO-PARAMETER WEIBULL DISTRIBUTION

We consider the two-parameter Weibull distribution with the shape parameter b and the scale parameter a. This distribution can be interpreted as a generalization of the exponential distribution, which is a special case as $b = 1$. The hazard rates of an exponential distribution are constant. The Weibull distribution might be used in order to model rising ($b > 1$) or falling ($b < 1$) hazard rates.

The density function of the two-parameter Weibull distribution is given by

$$f(x) = a \cdot b \cdot x^{b-1} \cdot exp(-a \cdot x^b) \quad a, b, x > 0.$$

Consequently,

$$F(x) = 1 - exp(-a \cdot x^b).$$

The problem of estimating the parameters of a Weibull distribution is described in detail by JOHNSON, KOTZ and BALAKRISHNAN (1994). Our estimation problem is estimating the parameters a and b, when the data are grouped. With given breaks $(\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_k)$ the cell counts n_i and the derived expressions are random variables. For the two-parameter Weibull we have: $\tilde{x}_0 = 0$ and $\tilde{x}_k = \infty$.

The different estimation methods are presented and compared in a simulation study in order to analyze the properties of the estimates for finite sample sizes.

4.1 MAXIMUM - LIKELIHOOD - ESTIMATION

The likelihoodfunction for the given case is:

$$\begin{aligned} L(n_1, n_2, \ldots, n_k; a, b) &= \prod_i \{P(\text{X in class i})(a, b)\}^{n_i} \\ &= \prod_i \{F(\tilde{x}_i; a, b) - F(\tilde{x}_{i-1}; a, b)\}^{n_i}. \\ &= \prod_i \left\{exp(-a \cdot \tilde{x}_{i-1}^b) - exp(-a \cdot \tilde{x}_i^b)\right\}^{n_i} \end{aligned} \tag{7}$$

The function $\ln(L(n_1, n_2, \ldots, n_k; a, b))$ is maximized with regard to the parameters a and b. There is no closed form for the parameter estimated and this function has to be maximize numerically.

4.2 MINIMUM - QUANTILE - DISTANCE - ESTIMATION

MQD estimation is a generalization of the ordinary quantile estimation, where all known quantiles are utilized for the parameter estimation. The quadratic form in equation 5 is minimized with regard to the unknown parameters a and b.

In case the weight matrix W is the unit matrix, equation 5 represents the sum of the squared distances of the empirical and the theoretical quantiles. This kind of estimation is called unweighted MQD estimation.

For the weighted MQD estimation the matrix $\mathbf{W}_3$ was used. If this matrix of weights is used, the parameter estimates can only be determined numerically. According to CARMODY, EUBANK and LARICCIA these estimates are unique, consistent and asymptotically normal.

Again, it has to be mentioned that the cell counts have to be interpreted as random variables, when the breaks are given. Therefore, in many applications, this method is not appropriate from a theoretical point of view.

4.3 OPTIMAL FIT OF THE CUMULATIVE DISTRIBUTION FUNCTION AFTER LINEARIZATION

The expression in equation 3 is minimized, where the inverse of the variance matrix of the vector with the cumulated relative frequencies, $V_{\hat{F}}$, is an appropriate matrix of weights.

For the two-parameter Weibull however, the cumulative distribution function can be linearized:

$$\hat{F}(\tilde{x}_i) = 1 - exp(-a\tilde{x}_i^b) \Leftrightarrow \ln(-\ln(1 - \hat{F}(\tilde{x}_i))) = \ln(a) + b \cdot \ln(\tilde{x}_i) \tag{8}$$

The problem of estimating the unknown parameters of a two-parameter Weibull distribution is hence reduced to fitting a weighted least squares equation.

$$\ln(-\ln(1-F_i)) = \beta_0 + \beta_1 \cdot \ln(\tilde{x}_i) + v_i \quad \text{for} \quad i = 1, \ldots, k-1 \tag{9}$$

For the determination of the variance matrix of the vector $\mathbf{v}$ we use the formula given in equation 6. With $h(\hat{F}) = \ln(-\ln(1-\hat{F}))$ it follows:

$$n \cdot \mathbf{V}_V = \begin{pmatrix} \frac{1-F_1}{F_1 \cdot \ln^2(1-F_1)} & \frac{1-F_2}{F_2 \cdot \ln(1-F_1) \cdot \ln(1-F_2)} & \cdots & \frac{1-F_{k-1}}{F_{k-1} \cdot \ln(1-F_1) \cdot \ln(1-F_{k-1})} \\ \frac{1-F_2}{F_2 \cdot \ln(1-F_1) \cdot \ln(1-F_2)} & \frac{1-F_2}{F_2 \cdot \ln^2(1-F_2)} & \cdots & \frac{1-F_2}{F_{k-1} \cdot \ln(1-F_2) \cdot \ln(1-F_{k-1})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1-F_{k-1}}{F_{k-1} \cdot \ln(1-F_1) \cdot \ln(1-F_{k-1})} & \frac{1-F_{k-1}}{F_{k-1} \cdot \ln(1-F_2) \cdot \ln(1-F_{k-1})} & \cdots & \frac{1-F_{k-1}}{F_{k-1} \cdot \ln^2(1-F_{k-1})} \end{pmatrix}.$$

On the basis of Equation 9 we can then perform a weighted least squares regression, where the inverse of $\mathbf{V}_V$ is an appropriate matrix of weights.

This is an Aitken-estimation of a least squares line. If the parameters of equation 9 are estimated OLS, the estimation procedure is called unweighted linearization method.

For the parameters of the Weibull-distribution it follows $\hat{b} = \hat{\beta}_1$ und $\hat{a} = exp(\hat{\beta}_0)$.

Additionally, the well-known results of Aitken-estimates may be carried over to the problem of estimating the shape parameter of a Weibull distribution.

4.4 DESCRIPTION OF THE SIMULATIONS

In order to analyze the properties of the alternative estimation methods simulations are performed. First, a random sample was drawn from a Weibull distribution with $a = 1$ and $b = 0.5, 1, 2$, respectively. Five classes were set up. The breaks were chosen to match the 0.2, 0.4, 0.6 and 0.8 quantile of the underlying theoretical distribution. Parameter estimation was then performed on the basis of the resulting frequency distributions. The statistical package S-PLUS was used to perform the simulations with 1000 replications. The S-PLUS function `nlmin()` was used to determine those estimators that do not have a closed form. In these cases the true parameter values were taken as starting values of the optimization algorithm. Therefore, no convergence problems were encountered. Additionally, it has to be mentioned that parameter b goes into the weight matrix proposed by CARMODY, EUBANK and LARICCIA (1984). Therefore, method MQDE performs slightly better than it would do in a real-world situation.

Method		Name
Maximum-Likelihood-Method		ML
Optimal fit of the cumulative distribution function after linearization	(unweighted)	LCDU
	(weighted)	LCD
Optimal fit of the cumulative distribution function	(unweighted)	OFU
	(weighted with F)	OFF
	(weighted with $\hat{F}$)	OFFhat
Minimum Quantile Distance Estimation	(unweighted)	MQDEU
	(weighted)	MQDE

Table 1: Estimation methods under consideration

The simulation results – mean, variance and mean square error (MSE) – are given in the appendix.

It is striking that almost all the methods lead to biased estimators. The bias disappears however, as the sample size increases. Biased, but asymptotically unbiased estimators will often yield the following typical pattern in simulation studies: The empirical biases have the same sign for different sample sizes, but the bias decreases as the sample size increases. For example we get for the maximum-likelihood-estimate of the parameter a:

true values of the parameters	mean estimate of a n=50	n=100	n=200
a=1, b=0.5	1.0217	1.0040	1.0033
a=1, b=1	1.0158	1.0052	1.0032
a=1, b=2	1.0200	1.0064	1.0035

Table 2: Bias of $\hat{a}$.

This pattern is a clear sign for the fact that the ML-estimator for the parameter a is – just like the ML-estimator for the parameter b – upward biased. The estimates resulting from weighted estimation of a straight line are upward biased with regard to parameter a and downward biased with regard to paramter b. The sizes of the biased are of similar magnitudes.

Another observation is the inferior influence of the bias on the MSE. The MSE has approximately the same value as the variance.

In order to compare the properties of the methods – analogous to the concept of relative efficiency – the MSEs of the different methods are compared to the MSE of the *best* method. Additionally, the ranks are given with the best method having rank 1.

	a=1, b=0.5											
	parameter a						parameter b					
	n=50		n=100		n=200		n=50		n=100		n=200	
LCDU	1.052	(4)	1.033	(4)	1.033	(5)	1.241	(6)	1.139	(6)	1.251	(7)
LCD	1	(1)	1	(1)	1.011	(2)	1.154	(4)	1.030	(2)	1.155	(6)
ML	1.030	(3)	1.013	(3)	1.019	(3)	1.067	(2)	1	(1)	1.098	(3)
MQDU	1.182	(8)	1.141	(5)	1.123	(8)	2.243	(8)	2.321	(8)	2.390	(8)
MQD	1.023	(2)	1.011	(2)	1.019	(4)	1.221	(5)	1.116	(4)	1.129	(4)
OFU	1.057	(5)	1.302	(7)	1.080	(7)	1	(1)	1.123	(5)	1.143	(5)
OFF	1.171	(7)	1.178	(6)	1	(1)	1.088	(3)	1.107	(3)	1.058	(2)
OFFhat	1.137	(6)	1.211	(8)	1.043	(6)	1.289	(7)	1.174	(7)	1	(1)

Table 3: Simulation results ($a = 1$, $b = 0.5$)

	a=1, b=1											
	parameter a						parameter b					
	n=50		n=100		n=200		n=50		n=100		n=200	
LCDU	1.043	(4)	1.066	(6)	1.114	(4)	1.169	(6)	1.080	(5)	1.109	(7)
LCD	1	(1)	1.032	(2)	1.108	(3)	1.136	(4)	1.048	(3)	1.033	(3)
ML	1.030	(3)	1.046	(4)	1.117	(7)	1	(1)	1	(1)	1	(1)
MQDU	1.053	(5)	1.067	(7)	1.115	(6)	1.301	(8)	1.322	(8)	1.286	(8)
MQD	1.009	(2)	1.040	(3)	1.114	(5)	1.066	(2)	1.008	(2)	1.019	(2)
OFU	1.119	(6)	1.057	(5)	1.105	(2)	1.147	(5)	1.073	(4)	1.056	(4)
OFF	1.148	(7)	1	(1)	1	(1)	1.097	(3)	1.117	(6)	1.085	(6)
OFFhat	1.204	(8)	1.166	(8)	1.155	(8)	1.260	(7)	1.142	(7)	1.064	(5)

Table 4: Simulation results ($a = 1$, $b = 1$)

	a=1, b=2											
	parameter a						parameter b					
	n=50		n=100		n=200		n=50		n=100		n=200	
LCDU	1.072	(6)	1.031	(6)	1.028	(6)	1.185	(8)	1.163	(8)	1.145	(8)
LCD	1.023	(2)	1	(1)	1.006	(2)	1.072	(3)	1.029	(5)	1.082	(6)
ML	1.055	(5)	1.014	(3)	1.012	(4)	1	(1)	1.006	(3)	1.054	(4)
MQDU	1.073	(7)	1.021	(5)	1.019	(5)	1.088	(6)	1.060	(6)	1.086	(7)
MQD	1.026	(3)	1.003	(2)	1.006	(3)	1.022	(2)	1.012	(4)	1.066	(5)
OFU	1	(1)	1.042	(8)	1.050	(7)	1.073	(4)	1.006	(2)	1.019	(3)
OFF	1.257	(8)	1.017	(4)	1	(1)	1.076	(5)	1	(1)	1.016	(2)
OFFhat	1.028	(4)	1.036	(7)	1.108	(8)	1.173	(7)	1.079	(7)	1	(1)

Table 5: Simulation results ($a = 1$, $b = 2$)

The interpretation of these simulation results is based on stable patterns within the simulation results as described above (table 2). However, in order to gain a feeling of how many digits can be believed in, we performed replicated simulations for the case $a = b = 1$ and $n = 200$. The results of these simulations are given in table A4.

With regard to parameter a method LCD dominates the ML-method in all the cases. With regard to parameter b the ML-method is superiour to the LCD method. In view of the fact that only the methods bases on the linearization of the cumulative distribution function (LCD and LCDU) lead to parameter estimates with a closed form and all other estimates have to be determined numerically, it can be concluded that method LCD represents a suitable estimation method. The properties of those estimates are comparable to the properties of the ML-estimates. On the other hand they require less calculation effort.

Therefore, the estimation method LCD can be recommended for real-world applications.

APPENDIX A1: SIMULATION RESULTS

		a=1, b=0.5					
		estimates for a			estimates for b		
		n=50	n=100	n=200	n=50	n=100	n=200
LCDU	Mean	1.0219	1.0049	1.0034	0.5162	0.5072	0.4995
	Variance	3.128	1.173	0.6425	0.8343	0.3452	0.1808
	MSE	3.172	1.174	0.643	0.8598	0.3464	0.1807
LCD	Mean	1.0268	1.0066	1.0047	0.4868	0.4927	0.4927
	Variance	2.947	1.134	0.6276	0.7827	0.3082	0.1618
	MSE	3.016	1.137	0.6292	0.7992	0.3132	0.1669
ML	Mean	1.0217	1.004	1.0033	0.5131	0.505	0.4988
	Variance	3.062	1.151	0.6339	0.7227	0.302	0.1586
	MSE	3.106	1.152	0.6343	0.7391	0.3042	0.1586
MQDU	Mean	1.0357	1.0111	1.0062	0.4996	0.4965	0.4959
	Variance	3.442	1.286	0.6959	1.556	0.7055	0.344
	MSE	3.566	1.297	0.699	1.554	0.706	0.3454
MQD	Mean	1.0152	1.0026	1.0021	0.5196	0.5073	0.5
	Variance	3.064	1.15	0.6366	0.847	0.3345	0.1631
	MSE	3.084	1.149	0.6346	0.8461	0.3394	0.1631
OFU	Mean	1.0262	1.0123	0.9989	0.511	0.5054	0.50384
	Variance	3.1225	1.4664	0.6728	0.6814	0.3391	0.1639
	MSE	3.188	1.48	0.6722	0.6927	0.3417	0.1652
OFF	Mean	1.0243	1.0101	1.0074	0.5179	0.5037	0.5031
	Variance	3.475	1.3303	0.6175	0.7426	0.3357	0.1521
	MSE	3.5305	1.339	0.6226	0.7535	0.3368	0.1529
OFFhat	Mean	1.0030	0.9995	0.9999	0.5118	0.4993	0.4984
	Variance	3.4319	1.379	0.6499	0.8936	0.3575	0.1444
	MSE	3.4294	1.3774	0.6492	0.8928	0.3572	0.1445

Table A1: Simulation results ($a = 1$, $b = 0.5$)

		a=1, b=1					
		estimates for a			estimates for b		
		n=50	n=100	n=200	n=50	n=100	n=200
LCDU	Mean	1.0164	1.0052	1.0034	1.0315	1.0127	1.0048
	Variance	2.687	1.387	0.6726	2.929	1.381	0.6883
	MSE	2.711	1.388	0.6731	3.026	1.396	0.6899
LCD	Mean	1.0214	1.008	1.0045	0.9692	0.9852	0.9909
	Variance	2.557	1.339	0.6678	2.848	1.333	0.6351
	MSE	2.6	1.344	0.6692	2.94	1.354	0.6427
ML	Mean	1.0158	1.0052	1.0032	1.2039	1.0108	1.0028
	Variance	2.655	1.361	0.6746	2.533	1.282	0.6219
	MSE	2.677	1.362	0.6749	2.588	1.292	0.622
MQDU	Mean	1.0222	1.0081	1.0048	1.0179	1.009	1.0008
	Variance	2.692	1.385	0.6719	3.337	1.702	0.8009
	MSE	2.739	1.39	0.6735	3.366	1.708	0.8002
MQD	Mean	1.0147	1.0051	1.0032	1.0055	1.0016	0.9984
	Variance	2.603	1.354	0.6726	2.758	1.303	0.6341
	MSE	2.623	1.355	0.673	2.758	1.302	0.6337
OFU	Mean	1.0211	1.0072	1.0037	1.0260	1.01566	1.00301
	Variance	2.867	1.3726	0.6667	2.904	1.3633	0.6565
	MSE	2.909	1.3765	0.66745	2.969	1.3865	0.6568
OFF	Mean	1.016	1.0067	0.9984	1.0279	1.01521	1.0104
	Variance	2.962	1.299	0.6044	2.7647	1.4213	0.665
	MSE	2.984	1.3024	0.6041	2.8398	1.443	0.6751
OFFhat	Mean	0.9951	0.9999	1.0022	0.9992	1.0063	0.9974
	Variance	3.1304	1.5196	0.69815	3.265	1.4731	0.6618
	MSE	3.1296	1.5181	0.69796	3.262	1.4756	0.6618

Table A2: Simulation results ($a = 1$, $b = 1$)

		a=1, b=2					
		estimates for a			estimates for b		
		n=50	n=100	n=200	n=50	n=100	n=200
LCDU	Mean	1.0201	1.0063	1.003	2.2569	2.0262	2.0123
	Variance	2.996	1.31	0.6401	12.53	5.829	2.871
	MSE	3.034	1.313	0.6403	12.84	5.892	2.883
LCD	Mean	1.0252	1.009	1.005	1.9412	1.9727	1.9859
	Variance	2.834	1.267	0.6246	11.27	5.145	2.705
	MSE	2.895	1.274	0.6264	11.61	5.214	2.722
ML	Mean	1.02	1.0064	1.0035	2.0439	2.0208	2.0124
	Variance	2.95	1.29	0.6298	10.65	5.056	2.639
	MSE	2.987	1.292	0.6304	10.83	5.094	2.652
MQDU	Mean	1.0242	1.0082	1.0045	2.075	2.0345	2.0206
	Variance	2.98	1.295	0.6334	11.24	5.245	2.694
	MSE	3.036	1.301	0.6348	11.79	5.37	2.734
MQD	Mean	1.0222	1.0077	1.0043	1.9746	1.988	1.9939
	Variance	2.857	1.273	0.626	11.01	5.116	2.681
	MSE	2.903	1.278	0.6267	11.07	5.125	2.682
OFU	Mean	1.01088	1.0085	1.0039	2.0461	2.0312	2.0106
	Variance	2.8211	1.322	0.6531	11.42	5.006	2.556
	MSE	2.8301	1.328	0.654	11.62	5.098	2.565
OFF	Mean	1.0136	1.0071	1.0064	2.0493	2.0179	2.0126
	Variance	3.543	1.292	0.6194	11.42	5.038	2.543
	MSE	3.558	1.296	0.6229	11.65	5.065	2.556
OFFhat	Mean	0.9995	0.9964	0.9974	2.0021	1.999	2.002
	Variance	2.914	1.3197	0.6904	12.73	5.471	2.519
	MSE	2.911	1.3196	0.6904	12.71	5.466	2.516

Table A3: Simulation results ($a = 1$, $b = 2$)

		$a = 1, b = 1, n = 200$					
		a	b			a	b
LCD		1.0045	0.9909	ML		1.0032	1.0028
	mean	1.0100	0.9925		mean	1.0088	1.0041
		1.0056	0.9935			1.0043	1.0054
		0.6678	0.6351			0.6746	0.6219
	var	0.6588	0.6229		var	0.6648	0.6163
		0.6564	0.6152			0.6609	0.6077

Table A4: Validity of simulation results

APPENDIX A2: S-PLUS-CODE

The following function migth be used to estimate the parameters of a Weibull distribtion with density

$$f(x) = a \cdot b \cdot x^{b-1} \cdot exp(-a \cdot x^b) \quad a, b, x > 0,$$

where *weibull.lcd[1]* is the estimator for a and *weibull.lcd[2]* is the estimator of b. The inputs x and y are vectors of equal length with **x**= $(\tilde{x}_1, \ldots, \tilde{x}_k)$ and **y**= $\hat{F}(\tilde{x}_1, \ldots, \hat{F}(\tilde{x}_k) =$ 1.

```
weibull.lcd<-function(x, y)
{
# This function finds the estimates a and b of a
# two-parameter weibull distribution.
```

```
# Inputs are:
# x: vector of quantiles (breaks)
# y: vector of corresponing proportions h(X<=x)

# initialization of sigma
  sigma <- matrix(rep(0, length(x)^2), nrow= length(x))
# data matrix
  X.stern <- matrix(c(rep(1, length(x)), log(x)), nrow = length(x), byrow = F)
# data matrix
    y.stern <- log( - log(1 - y))
# filling sigma:
 for(i in 1:length(x)) {
        for(j in 1:length(x))
            {sigma[i, j] <- (y[min(i, j)])/((1-
             y[min(i, j)]) * log(1 - y[i]) * log(1 - y[j]))
             }
                       }
# estimation of beta
        beta <- solve(t(X.stern) %*% solve(sigma) %*% X.stern) %*%
                t(X.stern) %*% solve(sigma) %*% y.stern
        return(c(exp(beta[1]), beta[2]))
}
```

REFERENCES

Carmody, T.J., Eubank, R.L., LaRiccia, V.N. (1984) A family of minimum quantile distance estimators for the three-parameter Weibull distribution. *Statistische Hefte*, Vol. 25, 69-82.

Cheng, K.F., Chen, C.H. (1988) Estimation of the Weibull parameters with grouped data. *Commun. Statist. - Theory Meth.*, Vol. 17(2), 325-341.

Cox, D.R. and Hinkley, D.V. (1974) Theoretical Statistics. Chapmann and Hall, London.

Gehan, E.A., Siddiqui, M.M. (1973) Simple regression methods for survival time studies. *Journal of the American Statistical Association*, Vol. 68(344), 848-856.

Johnson, N.L. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, Vol. 36, 149-176.

Johnson, N.L, Kotz, S., Balakrishnan N. (1994) Continuous univariate distributions. Vol. 1, 2nd Edition, Wiley, New York.

Lawless, J.F. (1982) Statistical models and methods for lifetime data. Wiley, New York.

McDonald, J.B., Ransom, M.R. (1979) Alternative parameter estimates based upon grouped data. *Commun. Statist. - Theory Meth.*, Vol. A8(9), 899-917.