Discussiones Mathematicae
Probability and Statistics 35 (2015) 75–94
doi:10.7151/dmps.1172

PROPERTIES OF THE GENERALIZED NONLINEAR LEAST SQUARES METHOD APPLIED FOR FITTING DISTRIBUTION TO DATA

Mirta Benšić

Department of Mathematics
University of Osijek
Trg Ljudevita Gaja 6
HR-31 000 Osijek, Croatia

e-mail: mirta@mathos.hr

Abstract

We introduce and analyze a class of estimators for distribution parameters based on the relationship between the distribution function and the empirical distribution function. This class includes the nonlinear least squares estimator and the weighted nonlinear least squares estimator which has been used in parameter estimation for lifetime data (see e.g. [6, 8]) as well as the generalized nonlinear least squares estimator proposed in [3]. Sufficient conditions for consistency and asymptotic normality are given. Capability and limitations are illustrated by simulations.

Keywords: generalized least squares, distribution fitting, generalized method of moments.

2010 Mathematics Subject Classification: 62F12.

1. Introduction

There is a huge number of methods which could be used for parameter estimation in the statistical model that describes trials of a random variable with the distribution unknown up to the parameter value. The right choice significantly depends on the distribution type as well as on the possible dependence between trials. If we focus on the specific distribution that is of interest for applications, a detailed investigation and comparison of different estimation methods are always welcome. For instance, such studies have been done for Weibull, generalized

Pareto, generalized Rayleigh, shifted Gompertz and many other distributions (see e.g. [1, 5, 8, 11, 14, 16, 18]). Here we introduce and analyze a class of estimators for distribution parameters based on the relationship between the distribution function and the empirical distribution function. We suggest to take this class into consideration while searching for the most appropriate estimator in the specified conditions and distribution type.

We primarily focus on the basic statistical model where a random sample $(X_1, \ldots, X_n)$ comes from independent trials of the random variable X with the distribution function $F(x;\theta)$ and $\theta \in \Theta \subseteq \mathbb{R}^p$ the unknown parameter, but generalization that allows a different dependence structure has also been discussed.

The class of estimators we introduce here includes the nonlinear least squares estimator and the weighted nonlinear least squares estimator which are used in parameter estimation (see e.g. [6, 8]) for lifetime data as well as the generalized nonlinear least squares estimator proposed in [3].

We have proved sufficient conditions for consistency and asymptotic normality for estimators which belong to the described class. Efficiency is also discussed. Quality that can be obtained with these estimators is illustrated by simulations. The nonlinear regression theory was the motivation to introduce these estimators, but it is not used in the proof of week consistency and asymptotic normality. For this we use the generalized method of moments (GMM). Namely, this method and properties of the distribution function allow us to assure consistency even without compactness of the parameter space.

Our presentation is organized as follows. In Section 2, the nonlinear regression framework is explained, which has been the motivation for the introduced class of estimators. Based on the GMM theory sufficient conditions for weak consistency are discussed in Section 3, while asymptotic normality and efficiency are presented in Section 4. It is important to point out that we do not need the assumption of marginal independence for the random sample in these considerations. In Section 5, we apply the presented theory to the independent and identically distributed (IID) random sample. Section 6 gives results of numerical experiments which illustrate capability of the proposed estimator class in the IID case as well as problems we sometimes face with these estimators. Results of an application to the real data set are given in Section 7. We conclude with some general remarks in Section 8.

2. Nonlinear generalized least squares framework

Let $\mathbf{x} = (x_1, \ldots, x_n)^\tau$ be a sample of n observations on a continuous random variable X with a distribution function $F(\cdot;\theta)$. By θ and Θ we denote an unknown finite-dimensional parameter to be estimated from the sample and the parame-

ter space, respectively. Throughout this section, $\hat{F}$ denotes a sample empirical distribution function, i.e.,

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(x_i).$$

The idea presented in [6] is to fit the distribution function $F(\cdot;\theta)$ to the data $(\hat{F}_n(x_1),\dots,\hat{F}_n(x_n))$ by the least squares method (LS), meaning to compute the parameter value $\hat{\theta}$ which minimizes the function

$$S_{LS}(\theta) = \sum_{i=1}^{n}(F(x_i,\theta) - \hat{F}_n(x_i))^2.$$

To discuss possible statistical reasons for choosing this estimator we consider the following nonlinear regression model for univariate responses $\hat{F}(x_i)$, $i \in \{1,\dots,n\}$:

$$\hat{F}_n(x_i) = F(x_i;\theta) + \varepsilon_i, \quad i \in \{1,\dots,n\}.$$

ε_i, $i \in \{1,\dots,n\}$ represent errors. At this moment, let us recall the empirical distribution function as a statistic. If we choose $x \in \mathbb{R}$ and an IID random sample $(X_1,\dots,X_n)$ from the distribution $F(\cdot;\theta)$, we know that the statistic

$$\frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

has an asymptotically normal distribution with mean $F(x;\theta)$[1]. This is the reason why we can rely on the fact that the errors have zero mean and a normal distribution for large n in the posted nonlinear regression model. This could be a classical case of nonlinear regression if the residuals were uncorrelated and homoscedastic, but they are not.

It is well-known in nonlinear regression theory with correlated errors that the covariance matrix of $(\varepsilon_1,\dots,\varepsilon_n)$, let us denote it by Σ, should be taken into account ([15]), so parameters are estimated by minimizing the function

$$S_{GLS}(\theta) = (\hat{F}_n(\mathbf{x}) - F(\mathbf{x};\theta))^\tau \Sigma^{-1}(\hat{F}_n(\mathbf{x}) - F(\mathbf{x};\theta)).$$

Here, $\mathbf{x}$ stands for $(x_1,\dots,x_n)^\tau$. For the real function f, $f(\mathbf{x})$ stands for $(f(x_1),\dots,f(x_n))^\tau$. This method is called the generalized least squares method (GLS).

Let us mention that both functions S_{LS} and S_{GLS} can be shown in the same form when using the appropriate $n \times n$ symmetric positive definite matrix A.

[1] The same is often true for random samples with identically distributed but not independent univariate margins (see e.g. [4]).

So, if we denote

$$S_A(\theta) = (\hat{F}_n(\mathbf{x}) - F(\mathbf{x};\theta))^\tau A(\hat{F}_n(\mathbf{x}) - F(\mathbf{x};\theta)),$$

then A is equal to the identity matrix I for the least square method and for the generalized least square method $A = \Sigma^{-1}$.

In this paper, we analyze properties of the estimator obtained by minimizing the function S_A, where A is a symmetric positive definite matrix. If the matrix A is applied, we will denote the estimator by $\hat{\theta}_A$.

Simulation results presented in [3] show that minimization of $S_{GLS} = S_{\Sigma^{-1}}$ is able to produce an estimator with good statistical properties, especially for relatively large or large data sets. By simulating IID samples from the Weibull distribution it is shown that this estimator is comparable to the maximum likelihood estimator by means of comparing the corresponding root mean square errors (RMSE).

Although the idea for the application of the least squares principle presented here comes from nonlinear regression theory, theoretical properties of the estimator cannot be taken automatically. Here we face a two-stage procedure. The first step is to compute empirical distribution from the data, i.e., to apply a nonparametric estimation method to estimate the distribution function. It is obvious that properties of the estimator applied in the first step influence properties of the unknown parameter estimator at the end. In the following sections we discuss statistical properties of the estimator $\hat{\theta}_A$ using the generalized method of moments (GMM) with appropriate moment condition.

3. Weak consistency

Let us assume that $\mathbf{x} = (x_1, \ldots, x_n)^\tau$ denotes the data coming from the random sample $\mathbf{X} = (X_1, \ldots, X_n)$ with identically distributed (not necessarily independent) univariate margins. Following the notification introduced in the previous section, we suppose that $F(\cdot;\theta_0)$ stands for the true common distribution function of univariate margins, which means that θ_0 is the true value of the unknown parameter $\theta \in \Theta$. Throughout this and the following sections we use the notation $\hat{F}_n(x)$ for the empirical distribution function.

To formulate the problem in a GMM setting (see e.g. [7]) let us choose $k \in \mathbb{N}$ arbitrary real numbers $u_1 < u_2 < \ldots < u_k$, $\mathbf{u} = (u_1 \ldots, u_k)^\tau$, satisfying

$$F(u_i;\theta_0) \neq F(u_j;\theta_0), \;\; i \neq j, \;\; F(u_1;\theta_0) \neq 0. \tag{1}$$

Based on these numbers we define the function $f : \mathbb{R} \times \Theta \to \mathbb{R}^k$ by

$$f(x,\theta) = (I_{(-\infty,u_1]}(x) - F(u_1;\theta), \ldots, I_{(-\infty,u_k]}(x) - F(u_k;\theta))^\tau$$

which will be used to set k moment conditions. Namely, it holds [2]

$$Ef(X_t, \theta_0) = 0, \ \forall t = 1, \dots, n.$$

By denoting $f_n(\theta)$ the corresponding sample moments:

$$f_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} f(X_t, \theta),$$

we are able to present the estimator $\hat{\theta}_A$ from the previous section as a GMM estimator. Namely, it holds that

$$f_n(\theta) = (\hat{F}_n(u_1) - F(u_1; \theta), \dots, \hat{F}_n(u_k) - F(u_k; \theta))^\tau.$$

If we choose a positive definite matrix A, we define a GMM estimator for θ:

$$\hat{\theta}_A(u_1, \dots, u_k) = \text{argmin}_\theta f_n(\theta)^\tau A f_n(\theta).$$

The estimator $\hat{\theta}_A$ from the previous section can be presented in this way if we choose $k = n$ and the ordered data set $(x_{(1)}, \dots, x_{(n)})$ for $\mathbf{u}$. In order to do this, the data $\mathbf{x}$ have to satisfy the condition

$$F(x_i; \theta_0) \neq F(x_j; \theta_0), \ \ i \neq j, \ F(x_{(1)}; \theta_0) \neq 0$$

(which is not restrictive for absolutely continuous random variables). Thus, $\hat{\theta}_A = \hat{\theta}_A(x_{(1)}, \dots, x_{(n)})$, meaning that the estimator

$$\hat{\theta}_A = \text{argmin}_\theta S_A(\theta)$$

can be presented in the form of a GMM estimator.

In the presented setting we have the GMM estimator for any selection of $k \in \mathbb{N}$ and $u_1, \dots, u_k \in \mathbb{R}$ satisfying condition (1). To ensure the base for identifiability of the parameter we must define a unique strategy of choosing $\mathbf{u}$. So, we have room to select appropriate k numbers $u_1, \dots, u_k$ to get desired estimator properties. This will be discussed in the sequel.

The results presented in [7] applied to the described model guarantee consistency of the GMM estimator if the following conditions are fulfilled:

C1) $g(\theta) = Ef(X_t, \theta)$ exists and it is finite for all $\theta \in \Theta$, $t = 1, \dots, n$.

C2) There exists a $\theta_0 \in \Theta$ such that $g(\theta) = 0$ if and only if $\theta = \theta_0$.

[2] As usual, when various operations on a function of $\mathbf{X} = (X_1, \dots, X_n)$ are taken, such as expectation or probability limit, the value θ_0 is used. Thus, we omit the subscript θ_0 from these notations.

C3) If $f_n(\theta) = ((f_n)_1(\theta), \dots, (f_n)_k(\theta))^\tau$, $g(\theta) = (g_1(\theta), \dots, g_k(\theta))^\tau$, then

$$\sup_{\theta\in\Theta} |(f_n)_j(\theta) - g_j(\theta)| \xrightarrow{P} 0, \quad \forall j = 1, \dots, k.$$

As $g_j(\theta) = F(u_j;\theta_0) - F(u_j;\theta)$, assumption C1 is fulfilled.

Assumption C3 is true if the random vector $\mathbf{X}$ satisfies conditions which guarantee that the empirical distribution function converges in probability to the true distribution function. Namely, $(f_n)_j(\theta) - g_j(\theta) = \hat{F}_n(u_j) - F(u_j;\theta_0)$, for all $j = 1, \dots, k$ and $\theta \in \Theta$, so that

$$\sup_{\theta\in\Theta} |(f_n)_j(\theta) - g_j(\theta)| = |\hat{F}_n(u_j) - F(u_j;\theta_0)|, \quad \forall j = 1, \dots, k.$$

To satisfy condition C2 we must be careful with the selection of $\mathbf{u}$. A reasonable way of choosing $\mathbf{u}$ is to use the data set. We use the maximal subset of $x_1, \dots, x_n$ such that assumptions (1) and C2 hold (if there is such a subset).

Assumption (1) is not restrictive for absolutely continuous distributions so that in almost all cases the whole data set can be used. Assumption C2 stated in terms of our moment condition means that

$$F(u_j;\theta_0) = F(u_j;\theta), \quad \forall j = 1, \dots, k \quad \Leftrightarrow \theta = \theta_0.$$

If there are no distribution functions from the desired class, $x \mapsto F(x;\theta_1)$ and $x \mapsto F(x;\theta_2)$ with intersections in all $u_1, \dots, u_k$ for $\theta_1 \neq \theta_2$, this assumption will be fulfilled. Thus, for k large enough and absolutely continuous distributions this assumption is also not too restrictive.

It is well-known from the theory of GMM estimators that the sequence of stochastic positive definite weighting matrices A_n can be used instead of the deterministic weighting matrix A. To guarantee consistency in this case, it is enough to ensure that

C4) there exists a non-random sequence of positive definite matrices $\bar{A}_n$ such that $A_n - \bar{A}_n \xrightarrow{P} 0$.

In the sequel, we use this property to discuss asymptotic efficiency as well as to choose the best estimator from the class in this sense.

4. Asymptotic efficiency

Assumptions in addition to those referring to consistency are needed to discuss asymptotic efficiency. We will use the assumptions presented in [7].

E1) $f(X_t, \theta)$ is continuously differentiable with respect to θ on Θ.

E2) For any sequence $\theta_n^\star$ such that $\theta_n^\star \xrightarrow{P} \theta_0$, $G_n(\theta_n^\star) - \bar{G}_n \xrightarrow{P} 0$.

Here $G_n(\theta) = \frac{\partial f_n(\theta)}{\partial \theta^\tau}$ and $\bar{G}_n$ is a sequence of matrices that do not depend on θ.

E3) $f(X_t, \theta_0)$ satisfies a central limit theorem, so that

$$\bar{V}_n^{-1/2}\sqrt{n} f_n(\theta_0) \xrightarrow{d} \mathbf{Z}, \;\; \mathbf{Z} \sim \mathcal{N}(0, I_k),$$

where $\bar{V}_n$ is a sequence of $k \times k$ non-random positive definite matrices defined as

$$\bar{V}_n = n \mathrm{Var} f_n(\theta_0).$$

Under assumptions C1–C4 and E1–E3 the estimator $\hat{\theta}_{A_n}(u_1, \ldots, u_k)$ is asymptotically normal and the asymptotic covariance matrix for a given sequence of weighting matrices A_n is

$$(\bar{G}_n^\tau \bar{A}_n \bar{G}_n)^{-1} \bar{G}_n^\tau \bar{A}_n \bar{V}_n \bar{A}_n \bar{G}_n (\bar{G}_n^\tau \bar{A}_n \bar{G}_n)^{-1}.$$

This result enable us to choose the weighting matrix to minimize the asymptotic covariance matrix.

It is well-known that the difference

$$(\bar{G}_n^\tau \bar{A}_n \bar{G}_n)^{-1} \bar{G}_n^\tau \bar{A}_n \bar{V}_n \bar{A}_n \bar{G}_n (\bar{G}_n^\tau \bar{A}_n \bar{G}_n)^{-1} - (\bar{G}_n^\tau \bar{V}_n^{-1} \bar{G}_n)^{-1}$$

is positive semi-definite for all $\bar{A}_n$. This suggests that, concerning efficiency, an inverse of the consistent estimator for $\bar{V}_n$ should be used as the weighting matrix (if it exists).

5. Independent and identically distributed random sample

Let us assume that $\mathbf{X} = (X_1, \ldots, X_n)$ is an IID random sample. In this case, properties of the empirical distribution function guarantee that assumptions C3 and E3 are fulfilled. Also, matrix sequences $\bar{A}_n$, $\bar{G}_n$ from assumptions C4 and E2 can be treated as constant matrices and the matrix $Var f_n(\theta_0)$ can be derived explicitly by using the true distribution function. Let us summarize the estimation procedure that ensures the consistent estimator and the way of choosing the weighting matrix to achieve the most efficient estimator for an IID sample within the described class of estimators.

To ensure consistency, it is enough to choose numbers $u_1, \ldots, u_k$ satisfying condition (1) and assumption C2. If the unknown distribution function is absolutely continuous and the data set large enough, it can be mainly done by choosing all different values from the data set.

If the function $\theta \mapsto F(x;\theta)$ is continuously differentiable on Θ, for all $x \in \mathbb{R}$, then assumption E1 is satisfied. Also, continuity of the derivative in this case ensures that whenever we have a sequence $\theta_n^\star$ converging to θ_0 in probability,

$$G_n(\theta_n^\star) = G(\theta_n^\star) = -\frac{\partial F(\mathbf{u};\theta)}{\partial \theta^\tau}|_{\theta_n^\star} \xrightarrow{P} -\frac{\partial F(\mathbf{u};\theta)}{\partial \theta^\tau}|_{\theta_0}$$

meaning that assumption E2 is also satisfied. Thus, to achieve asymptotic normality of the estimator $\hat{\theta}_A(u_1, \ldots, u_k)$, in addition to condition (1) and assumption C2, the sufficient condition is that the function $\theta \mapsto F(x;\theta)$ is continuously differentiable on Θ for all $x \in \mathbb{R}$.

In this case, an inverse of the matrix

$$V_k = \mathrm{Var}[(I_{(-\infty,u_1]}(X_t) - F(u_1;\theta_0), \ldots, I_{(-\infty,u_k]}(X_t) - F(u_k;\theta_0))^\tau]$$

should be chosen as the weighting matrix to achieve the most efficient estimator from the class of estimators $\{\hat{\theta}_A(u_1, \ldots, u_k) : A \text{ positive semidefinite}\}$. The value θ_0 is unknown but GMM theory allows us to use the consistent estimator instead (see e.g. [7]). If we compute the matrix V_k we can see that it has a nice form:[3]

$$V_k = F(u_i \wedge u_j;\theta_0) - F(u_i;\theta_0)F(u_j;\theta_0), \quad i,j \in \{1, \ldots, k\}. \tag{2}$$

At the end of this section, we conclude that the estimator $\hat{\theta}_A(u_1, \ldots, u_k)$ is consistent with the right choice of $(u_1, \ldots, u_k)$ for each positive definite non-stochastic matrix A. In the case of a continuously differentiable distribution function with respect to the parameter, the estimator will also be asymptotically normal with the asymptotic covariance matrix that can be easily calculated.

6. Numerical experiments and discussion

In this section, we present results of some numerical experiments to compare the performance of the introduced class of estimators and other estimators with respect to their observed biases and mean squared errors (MSE). We compute the average estimates, observed MSE and bias over replications of IID samples for different sample sizes. All numerical computations were carried out with R. Optimizations in all cases were done by the procedures `optim()` or `nlm()` depending on the function to be optimized.

[3] $t_i \wedge t_j = min\{t_i, t_j\}$.

For the purpose of comparison with $\hat{\theta}_A$, we choose the maximum likelihood estimator and two typical estimators based on the empirical distribution used in the literature: the goodness-of-fit estimators with the Cramer-von Mises statistic and with the Anderson-Darling statistic ([11, 13]).

6.1. Weighting matrices

For simulations we have chosen three different forms of the weighting matrix: I, $\hat{V}_n^{-1}$ and W_n^{-1}. Here, I is an identity matrix, $\hat{V}_n^{-1}$ stands for the inverse of the matrix V_n defined in (2) and evaluated in the consistent estimator of θ. W_n^{-1} is an inverse of the diagonal matrix with diagonal elements which consistently estimate the diagonal elements from V_n. The case with $A = I$ corresponds to the LS estimator from Section 2, so we call this estimator the least squares estimator and denote it by $\hat{\theta}_{LS}$. If we use the matrix $\hat{V}_n^{-1}$, we get the estimator that corresponds to the generalized least squares approach from Section 2, so we call it the generalized least square estimator and denote it by $\hat{\theta}_{GLS}$. The case with the matrix W_n^{-1} corresponds to the method usually called the weighted least squares so we denote this estimator by $\hat{\theta}_{WLS}$.

We simulated random samples only for absolutely continuous distributions and used the ordered sample as $(u_1, \ldots, u_n)$. To be able to discuss efficiency we chose only those distributions that satisfy assumption E1. As the LS estimator was consistent for our data, we use LS estimation to compute matrices $\hat{V}_n^{-1}$ and W_n^{-1} for GLS and WLS estimations.

6.2. Exponential distribution

To illustrate a quality level we can get by means of the introduced estimator $\hat{\theta}_A$ and a wise choice of the matrix A we simulated random samples from the exponential distribution, $F(x;\theta) = 1 - e^{-\frac{x}{\theta}}$, $x > 0$. It is known for this distribution that the maximum likelihood estimator is the sample mean ($\hat{\theta}_{ML} = \bar{X}_n$), it is unbiased and its variance attains the Cramér-Rao lower bound (it is fully efficient), i.e., the best we can expect for the estimator in the mean square error sense. We found out that the value of the parameter does not influence the rating of estimators significantly, so we present and discuss only the results for $\theta = 2$.

For the exponential distribution the weighting matrices for GLS and WLS estimation are computed as the inverse of

$$V_n = \left[1 - e^{-\frac{x_i \wedge x_j}{\theta}} - (1 - e^{-\frac{x_i}{\theta}})(1 - e^{-\frac{x_j}{\theta}})\right]_{i,j}, \quad i, j = 1, \ldots, n,$$

or

$$W_n = \mathrm{diag}(1 - e^{-\frac{x_1}{\theta}} - (1 - e^{-\frac{x_1}{\theta}})^2, \ldots, 1 - e^{-\frac{x_n}{\theta}} - (1 - e^{-\frac{x_n}{\theta}})^2),$$

respectively, evaluated at LS estimation.

To compute estimates for the asymptotic covariance matrix of the LS, WLS and GLS estimators we used the expression:

$$\operatorname{Var}\hat{\theta}_{A_n} = (G_n^\tau A_n G_n)^{-1} G_n^\tau A_n V_n A_n G_n (G_n^\tau A_n G_n)^{-1},$$

where A_n stands for the inverse of the chosen matrix (I, W_n or V_n) evaluated at the estimated parameter value,

$$G_n = \frac{1}{\theta^2} \begin{bmatrix} x_1 e^{-\frac{x_1}{\theta}} \\ \vdots \\ x_n e^{-\frac{x_n}{\theta}} \end{bmatrix}$$

(also evaluated at the estimated parameter value). For the LS and GLS estimators this expression can be simplified so we used expressions

$$\operatorname{Var}\hat{\theta}_{LS} = (G_n^\tau G_n)^{-1} G_n^\tau V_n G_n (G_n^\tau G_n)^{-1},$$

and

$$\operatorname{Var}\hat{\theta}_{GLS} = (G_n^\tau V_n G_n)^{-1}.$$

Table 1 presents average estimates for asymptotic variances of $\hat{\theta}_{LS}$, $\hat{\theta}_{WLS}$ and $\hat{\theta}_{GLS}$ over 1000 replications as well as the ML variances for different sample sizes. Figure 8 shows these values relative to the $\hat{\theta}_{ML}$ variance for different sample sizes.

The results confirm our expectation founded on the theory, i.e. the values for the LS and WLS variances are always greater than the values for the GLS variances regardless of the sample dimension. Besides, the values for the GLS variances are really close to the ML variances especially for large sample sizes.

The observed mean square errors and biases over 1000 replications are shown in Table 2 for all estimators. Figure 8 presents the observed MSE relative to the observed MSE of the ML estimator and biases for all estimators. Regarding MSE we can notice similarities in behavior for pairs of estimators: LS and CM, WLS and AD, GLS and ML, especially for large sample sizes. It is important to point out that the GLS estimator attained MSE level of the best possible estimator for the assumed distribution and large sample sizes.

6.3. Generalized Rayleigh distribution

J.G. Surles, W.J. Padgett in [17] proposed the two-parameter Burr Type X distribution:

$$F(x;\alpha,\lambda) = (1 - e^{-(\lambda x)^2})^\alpha, \qquad x > 0,\ \alpha > 0,\ \lambda > 0$$

and named it the generalized Rayleigh distribution. They observed that it can be used quite effectively in modeling strength data and also general lifetime data.

Different properties of the generalized Rayleigh distribution are summarized in [9]. For $\alpha \leq 0.5$ it has a decreasing density function. Otherwise, its density is unimodal.

In [8], different estimation procedures are used to estimate the unknown parameters and their performances are compared by simulations. It is shown that, regarding MSE, the ML estimator can be outperformed by other estimators. Here we compare the behavior of LS, WLS, GLS, AD, CM and ML estimators applied to this distribution. An average estimate, the observed MSE and bias are computed over 500 replications for different sample sizes. The parameter λ is a scale parameter and its value does not influence the rating among estimators significantly but the value of the shape parameter α does. So, we present results for $\lambda = 1$ and different shape values. Average estimates for theoretical variances of LS, WLS and GLS estimators are also computed and compared with the theoretical variance of the ML estimator.

All matrices and estimations are computed analogously to the procedure presented in Subsection 6.2. As a starting point for the ML and LS estimations we used $\alpha_0 = 0.5$ for decreasing densities and $\alpha = 1$ for unimodal densities. The value for λ_0 was then calculated from the equation

$$\text{median}(\alpha, \lambda) = \left[-\frac{1}{\lambda} \ln(1 - (\frac{1}{2})^{\frac{1}{\alpha}}) \right]^{\frac{1}{2}},$$

evaluated at the sample median. As a starting point for the WLS and GLS estimations we used $\hat{\theta}_{LS}$. We believe this is reasonable as the same value is used for calculation of the weighting matrices for these methods.

Average estimates of asymptotic variances and different shape values for $n = 30$ and $n = 500$ are presented in Table 3 and in Table 4, respectively. Figure 8 shows average estimates of asymptotic variances computed relatively to the maximum likelihood variance for $n = 500$. These results confirm our expectations. The GLS estimator has the smallest variance in comparison with LS and WLS estimators but greater than the ML asymptotic variance for all shape values. In addition to that, values of the asymptotic variances for the GLS estimator are really close to the values of the asymptotic variances for the ML estimator if the sample size is large enough.

Average parameter estimates and the observed MSEs for $n = 30$ and different shape values are presented in Table 5. For $n = 500$, results are summarized in Table 6. Figure 8 shows average observed bias calculated relatively to the true parameter value for $n = 500$ and Figure 8 shows the average observed MSE calculated relatively to the true parameter value for different shape values. Here we can also notice similarities in the behavior of LS and CM estimators. Some similarities in the behavior of WLS and AD estimators are present in most cases on both sample sizes although not as consistently as for the exponential distri-

bution. With respect to the pair of ML and GLS estimators we can notice, as for the exponential distribution, the different behavior in bias. While the ML estimator mostly overestimates parameter values, the GLS estimator mostly underestimates them with the exception of the estimation of λ if the α values are not large.

Concerning MSE, for small sample sizes WLS estimators outperform the ML estimator for α and in these cases the GLS mostly ranks second, i.e., it also mostly outperforms the MLE. However, for the estimation of λ, the MLE always has the smallest MSE. For large sample sizes, the MLE is always the best. What is important to notice here is that the MSE of the GLS estimator are not similar to the MSE of the ML estimator for large sample sizes. The WLS and AD estimators are better although the asymptotic variance of the GLS estimator is really small and close to the asymptotic variance of the ML estimator. It was shown in [3] that the similarities in MSEs between ML and GLS estimators are also present if the model function is the two-parameter Weibull and in some cases for the three-parameter Weibull model. It means that an increase in the parameter dimension cannot be the only reason for the aforementioned difference between GLS and ML mean square errors in the generalized Rayleigh model. While investigating why this happened, we found out that the differences in biases are one part of the problem but not the only one. The observed variances for the GLS estimator in this model are often greater than the theoretical variances. We have noticed the appearance of outliers in the set of GLS estimations caused by the fact that the GLS procedure has been most sensible to initial conditions.

7. Real data example

In this section, we apply the described estimation procedures on real data. In [12], the data on failure and service times for a particular large aircraft model windshield are given. We fit the generalized Rayleigh distribution on the complete part of data, i.e., failure time data only. The obtained parameter values and the corresponding values of the Kolmogorov-Smirnov goodness of fit test statistics (GOF_{KS}) are shown in Table 7. As it can be seen, all models are acceptable and the GLS estimation gives the smallest value of the Kolmogorov-Smirnov statistic.

8. Concluding remarks

It is not rare to find the LS or WLS method for parameter estimation in applications if the data come from independent observations on a random variable with a parametric distribution function. In doing so, an expression for the distribution function in its explicit or suitably transformed form is used as the model function

and the empirical distribution (or its transformation which corresponds to the transformation of applied distribution function) as the response. Transformations are usually applied in order to linearize the model function in parameters to ease the calculation in optimization procedures. This paper introduces and analyzes the class of estimators that generalizes the mentioned LS and WLS approach when using the distribution function in its original form (mostly nonlinear in the parameters).

In this study sufficient conditions are given which assure consistency of the estimator if it belongs to the introduced class. The most efficient estimator from this class has also been identified. While analyzing consistency and asymptotic normality the GMM method has been applied with appropriate moment condition.

Our study confirms that the suggested class is reasonable for applications. The conditions that are sufficient for consistency in the IID random sample can be fulfilled easily. The exponential distribution case confirms that the type of the weighting matrix can be chosen such that the observed MSE is comparable to the observed MSE of the best possible estimator. However, the application to the generalized Rayleigh distribution illustrates that special attention should be paid to the sensitivity of numerical procedures if non-diagonal weighting matrices are used. It is also worth mentioning that computations of these estimators can be carried out with any software for nonlinear regression.

In addition, it is also important that the same class of estimators can be applied if we deal with random samples with identically distributed but not independent univariate margins in the case when the empirical distribution function is asymptotically normal. In that case, in order to discuss the efficiency of different estimators from the introduced class, the asymptotic covariance matrix of the empirical distribution function should be calculated and estimated consistently.

References

[1] P.A. Al-Baidhani and C.D. Sinclair, *Comparison of methods of estimation of parameters of the Weibull distribution*, Communications in Statistics – Simulation and Computation **16** (1987) 373–384.

[2] T.W. Anderson and D.A. Darling, *Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes*, Ann. Math. Statist. **23** (1952) 193–212.

[3] M. Benšić, *Fitting distribution to data by a generalized nonlinear least squares method*, Communications in Statistics – Simulation and Computation (2012), in press.

[4] P. Billingsly, Convergence of Probability Measures (John Wiley & Sons, New York, 1968).

[5] J. Castillo and J. Daoudi, *Estimation of the generalized Pareto distribution*, Statistics and Probability Letters **79** (2009) 684–688.

[6] D. Jukić, M. Benšić and R. Scitovski, *On the existence of the nonlinear weighted least squares estimate for a three-parameter Weibull distribution*, Computational Statistics & Data Analysis **52** (2008) 4502–4511.

[7] D. Harris and L. Matyas, *Introduction to the generalized method of moment estimation*, in: Matyas, L. (Ed.), Generalized method of moment estimation (Cambridge University Press, Cambridge, 1999) 3–30.

[8] D. Kundu and M.Z. Raqab, *Generalized Rayleigh distribution: different methods of estimations*, Computational Statistics & Data Analysis **49** (2005) 187–200.

[9] D. Kundu and M.Z. Raqab, *Burr Type X distribution: Revisited*, J. Prob. Stat. Sci. **4** (2006) 179–193.

[10] J.F. Lawless, Statistical Models and Methods for Lifetime Data (Wiley, New York, 1982).

[11] A. Luceño, *Fitting the generalized Pareto distribution to data using maximum goodness-of-fit estimator*, Comp. Stat. & Data Anal. **51** (2006) 904–917.

[12] D.N.P. Murthy, M. Bulmer and J.A. Eccleston, *Weibull model selection for reliability modelling*, Reliability Engineering and System Safety **86** (2004) 257–267.

[13] D. Pollard, *The minimum distance method of testing*, Metrika **27** (1980) 43–70.

[14] H. Rinne, The Weibull Distribution. A Handbook (Chapman & Hall/CRC, Boca Raton, 2009).

[15] G.A.F. Seber and C.J. Wild, Nonlinear Regression (John Wiley & Sons, New York, 1989).

[16] R.L. Smith and J.C. Naylor, *A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution*, Biometrika **73** (1987) 67–90.

[17] J.G. Surles and W.J. Padgett, *Inference for reliability and stress-strength for a scaled Burr Type X distribution*, Lifetime Data Analysis **7** (2001) 187–200.

[18] F.J. Torres, *Estimation of parameters of the shifted Gompertz distribution using least squares, maximum likelihood and moments methods*, J. Comp. and Appl. Math. **255** (2014) 867–877.

[19] J. Wolfowitz, *Estimation by minimum distance method*, Ann. Inst. Statisti. Math. **5** (1953) 9–23.

[20] J. Wolfowitz, *The minimum distance method*, Ann. Math. Statist. **28** (1957) 75–88.

Received 20 April 2015

APPENDIX: TABLES AND FIGURES

Table 1. Average estimates of asymptotic variances for exponential distribution, $\theta = 2$, and different sample sizes.

n	30	50	200	300	1000
LS	0.423897	0.326620	0.162420	0.132503	0.072501
WLS	0.402877	0.309615	0.153704	0.125341	0.068535
GLS	0.374088	0.286837	0.141939	0.115734	0.063292
ML	0.365148	0.282843	0.141421	0.115470	0.063246

Table 2. Average estimates and observed mean square errors (in parentheses) for exponential distribution, $\theta = 2$.

n	30	50	200	300	1000
LS	1.907124	1.955387	1.985529	1.992158	2.002336
	(0.183494)	(0.104294)	(0.027263)	(0.018308)	(0.004848)
CM	2.017626	2.022541	2.002338	2.003403	2.005720
	(0.192449)	(0.108910)	(0.027447)	(0.018429)	(0.004889)
WLS	1.932000	1.965325	1.985404	1.992233	2.002145
	(0.169038)	(0.100557)	(0.024876)	(0.016240)	(0.004386)
AD	2.026392	2.027167	2.002878	2.004482	2.006043
	(0.171898)	(0.099050)	(0.024659)	(0.016412)	(0.004434)
GLS	1.882197	1.930812	1.973390	1.985531	1.997067
	(0.148162)	(0.086543)	(0.022504)	(0.013802)	(0.003868)
ML	1.990163	2.007764	1.996572	2.002243	2.003335
	(0.142825)	(0.085568)	(0.021549)	(0.013746)	(0.003889)

Table 3. Average estimates of asymptotic variances for Generalized Rayleigh distribution, $n = 30$, $\lambda = 1$, and different shape values.

α	0.4	0.8	1.5	2	3	10
LS α	0.010059	0.050017	0.230282	0.471070	1.286170	25.780385
LS λ	0.042271	0.024703	0.017323	0.015325	0.013219	0.009945
WLS α	0.008487	0.041232	0.187216	0.379014	1.020767	20.294971
WLS λ	0.033637	0.020151	0.014266	0.012626	0.010874	0.008216
GLS α	0.007751	0.037580	0.169003	0.338337	0.907569	17.915822
GLS λ	0.029423	0.018123	0.013013	0.011485	0.009871	0.007441
ML α	0.007032	0.033911	0.149267	0.298303	0.800000	15.463000
ML λ	0.025476	0.016173	0.011664	0.010311	0.008889	0.006601

Table 4. Average estimates of asymptotic variances for Generalized Rayleigh distribution, $n = 500$, $\lambda = 1$, and different shape values.

α	0.4	0.8	1.5	2	3	10
LS α	0.000574	0.002881	0.013159	0.026752	0.073053	1.469135
LS λ	0.002392	0.001409	0.000994	0.000877	0.000756	0.000573
WLS α	0.000475	0.002322	0.010403	0.020908	0.056628	1.116500
WLS λ	0.001846	0.001121	0.000800	0.000705	0.000609	0.000458
GLS α	0.000424	0.002048	0.009025	0.018043	0.048449	0.937636
GLS λ	0.001543	0.000977	0.000705	0.000623	0.000537	0.000399
ML α	0.000422	0.002035	0.008956	0.017898	0.048000	0.927780
ML λ	0.001529	0.000970	0.000700	0.000619	0.000533	0.000396

Table 5. Average estimates and observed mean square errors (in parentheses) for Generalized Rayleigh distribution, $n = 30$, $\lambda = 1$, and different shape values.

α	0.4	0.8	1.5	2	3	10
LS α	0.432893	0.860601	1.661423	2.254612	3.272273	13.484353
	(0.013400)	(0.091111)	(0.359988)	(0.834189)	(1.540460)	(130.513337)
LS λ	1.085505	1.061082	1.038937	1.033296	1.024192	1.023652
	(0.058880)	(0.031376)	(0.019265)	(0.018375)	(0.012752)	(0.010354)
CM α	0.447678	0.888288	1.706996	2.309987	3.349705	13.654724
	(0.015275)	(0.101462)	(0.391572)	(0.900606)	(1.669251)	(130.038509)
CM λ	1.061075	1.044495	1.026292	1.021452	1.013784	1.015902
	(0.052854)	(0.029071)	(0.018233)	(0.017443)	(0.012341)	(0.010005)
WLS α	0.398263	0.763321	1.463683	1.938331	2.810385	9.851301
	(0.007721)	(0.044918)	(0.213034)	(0.400237)	(0.989017)	(23.594586)
WLS λ	1.037518	1.013931	1.000923	0.995073	0.988803	0.987617
	(0.040573)	(0.023266)	(0.015297)	(0.013576)	(0.011125)	(0.008305)
AD α	0.429988	0.832667	1.585894	2.136508	3.110749	11.755775
	(0.009185)	(0.054399)	(0.231389)	(0.479958)	(1.056463)	(49.394037)
AD λ	1.063905	1.015844	1.002931	1.001582	0.997300	1.002945
	(0.154878)	(0.020692)	(0.013617)	(0.012704)	(0.009536)	(0.007615)
GLS α	0.392644	0.747209	1.447103	1.918643	2.776443	9.855416
	(0.013907)	(0.048398)	(0.221782)	(0.441611)	(1.089020)	(25.995926)
GLS λ	1.069338	1.031284	1.013151	1.004374	0.996518	0.995297
	(0.041786)	(0.022713)	(0.014780)	(0.013910)	(0.011034)	(0.008826)
ML α	0.443742	0.866889	1.682073	2.263892	3.322782	12.885790
	(0.010477)	(0.056855)	(0.268958)	(0.560751)	(1.297896)	(56.228781)
ML λ	1.065402	1.041235	1.027712	1.022436	1.017886	1.024105
	(0.035947)	(0.020530)	(0.013476)	(0.012486)	(0.009637)	(0.008165)

Table 6. Average estimates and observed mean square errors (in parentheses) for Generalized Rayleigh distribution, $n = 500$, $\lambda = 1$, and different shape values.

α	0.4	0.8	1.5	2	3	10
LS α	0.401072	0.804120	1.509713	2.006822	3.004879	10.234033
	(0.000609)	(0.002970)	(0.012460)	(0.026548)	(0.070820)	(1.696478)
LS λ	1.005191	1.004089	1.005009	1.001493	1.000323	1.003419
	(0.002630)	(0.001396)	(0.000935)	(0.000841)	(0.000703)	(0.000577)
CM α	0.401863	0.805580	1.512148	2.009860	3.008891	10.242140
	(0.000613)	(0.002994)	(0.012554)	(0.026676)	(0.071049)	(1.702099)
CM λ	1.003750	1.003114	1.004251	1.000808	0.999704	1.002931
	(0.002611)	(0.001387)	(0.000928)	(0.000839)	(0.000703)	(0.000574)
WLS α	0.399876	0.800722	1.504045	1.996725	2.989459	10.091798
	(0.000491)	(0.002344)	(0.009690)	(0.021148)	(0.060599)	(1.187263)
WLS λ	1.003179	1.002238	1.003779	1.000209	0.999024	1.001536
	(0.002085)	(0.001083)	(0.000765)	(0.000694)	(0.000725)	(0.000443)
AD α	0.401026	0.803076	1.508705	2.002868	3.003136	10.132350
	(0.000498)	(0.002378)	(0.009846)	(0.021437)	(0.054144)	(1.223646)
ADλ	1.001782	1.001422	1.003280	0.999769	0.999285	1.001464
	(0.002078)	(0.001083)	(0.000761)	(0.000696)	(0.000562)	(0.000448)
GLS α	0.398127	0.795783	1.488681	1.977198	2.955059	9.758585
	(0.000440)	(0.002399)	(0.011740)	(0.022673)	(0.075353)	(1.430501)
GLS λ	1.005308	1.001904	1.001571	0.998729	0.997061	0.996265
	(0.001783)	(0.001040)	(0.000905)	(0.000746)	(0.000962)	(0.000591)
ML α	0.402126	0.805212	1.515311	2.009295	3.025369	10.142795
	(0.000441)	(0.002091)	(0.009122)	(0.019694)	(0.048308)	(1.044092)
ML λ	1.004185	1.002522	1.004530	1.000894	1.001474	1.002106
	(0.001752)	(0.000934)	(0.000743)	(0.000639)	(0.000514)	(0.000388)

Table 7. Estimates and the corresponding values of the Kolmogorov-Smirnov goodness of fit test statistic for the failure time data set and generalized Rayleigh distribution.

method	$\hat{\alpha}$	$\hat{\lambda}$	GOF_{KS}
ML	1.200811	0.376297	0.060685
LS	1.380761	0.389711	0.049944
WLS	1.305006	0.381751	0.052284
GLS	1.379113	0.392855	0.045706
CM	1.388085	0.387328	0.056582
AD	1.362369	0.384189	0.055938

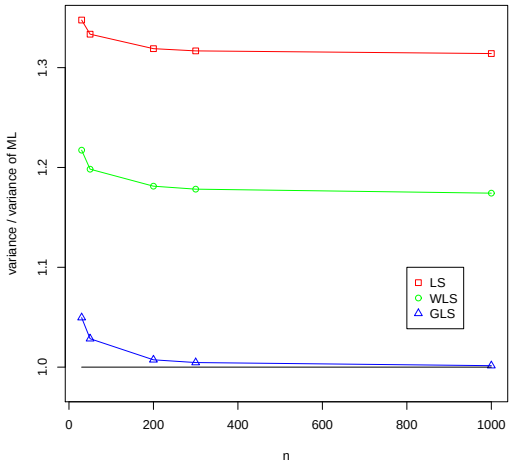

Figure 1. Average estimates of asymptotic variance for $\hat{\theta}_{LS}$, $\hat{\theta}_{WLS}$ and $\hat{\theta}_{GLS}$ relative to $\hat{\theta}_{ML}$ variance (Exponential distribution, $\theta = 2$).

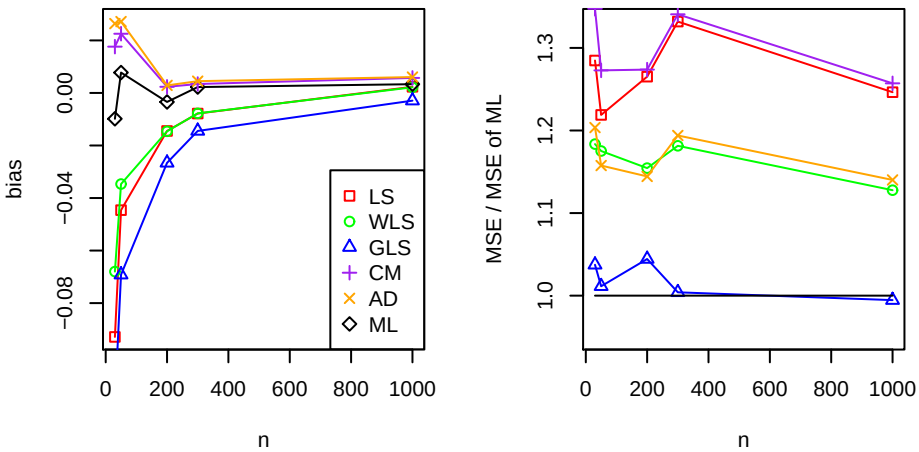

Figure 2. Observed bias and observed MSE relative to the observed maximum likelihood MSE for exponential distribution, $\theta = 2$.

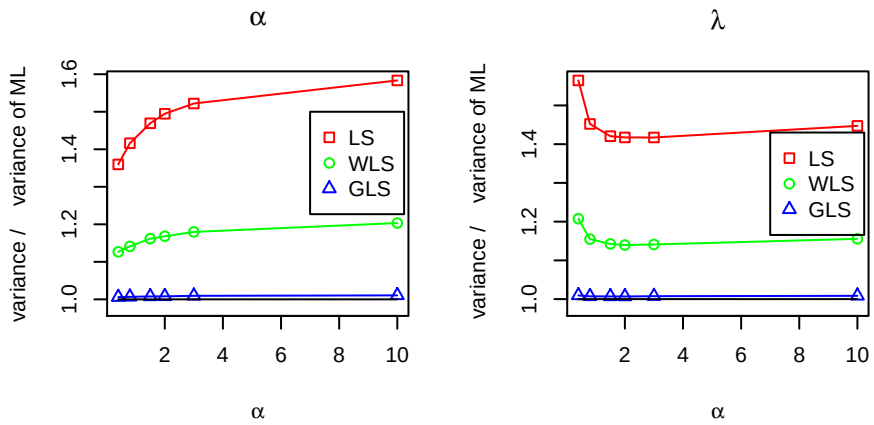

Figure 3. Average estimates of asymptotic variance relative to the maximum likelihood variance for different shape values (Generalized Rayleigh distribution, $\lambda = 1$, n=500) for LS, WLS and GLS estimation procedures.

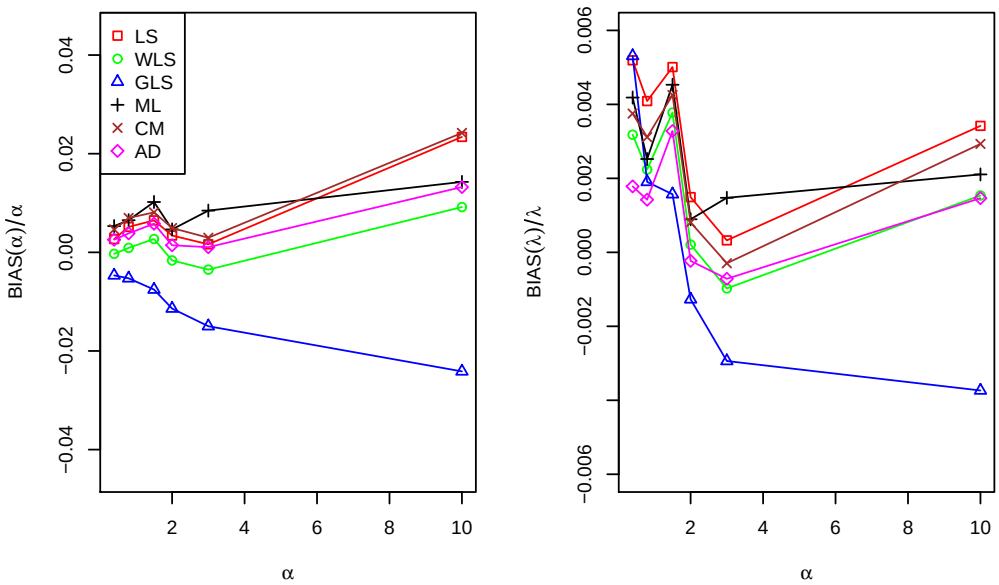

Figure 4. Average observed bias relative to the true parameter value for different shape values (Generalized Rayleigh distribution, $\lambda = 1$, n=500) for LS, WLS, GLS and ML estimation procedures.

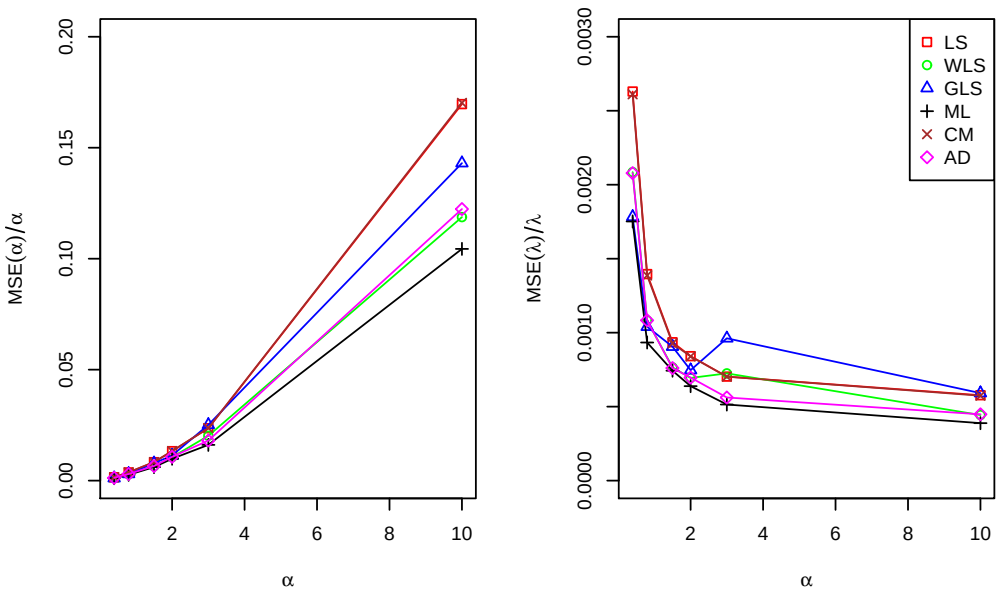

Figure 5. Average observed MSE relative to the true parameter value for different shape values (Generalized Rayleigh distribution, $\lambda = 1$, n=500) for LS, WLS, GLS and ML estimation procedures.