VILNIUS UNIVERSITY

FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS MASTER'S STUDY PROGRAMME

Master's thesis

Multiple Outliers Identification Method in Accelerated Failure Time (AFT) – Regression Models

Daugybinių išskirčių identifikavimo metodas pagreitinto gedimų laiko (AFT) - regresijos modeliams

Deivydas Sinkevičius

Supervisor Prof., Habil. dr. Vilijandas Bagdonavičius

VILNIUS 2021

Daugybinių išskirčių identifikavimo metodas pagreitinto gedimų laiko (AFT) - regresijos modeliams

Santrauka

Šio tyrimo tikslas pateikti išskirčių ieškojimo metodų normaliosioje tiesinėje regresijoje modifikacijas pagreitinto gedimų laiko (AFT) regresiniams modeliams ir palyginti juos. Šiame darbe nagrinėjami trys išskirčių identifikavimo metodai: BP ir David-Gather (DG), grindžiamu vienu iš dviejų vertinimo metodų: mažiausiųjų kvadratų bei robastiniu. Pirmosios dalies tikslas modifikuoti BP išskirčių ieškojimo metodą pagreitinto gedimų laiko (AFT) regresiniams modeliams. Antrosios dalies tikslas: generuojant duomenis palyginti visus tris metodus bei pateikti praktinius pavyzdžius. Naudojant duomenų generavimą gauta, kad daugelyje situacijų BP metodas geriau identifikuoja išskirtis už abu DG metodus. Atlikus procedūrą su realiais duomenimis, rezultatai gauti tokie patys, kaip ir simuliacijų metu. Gauti rezultatai įrodo, kad BP metodą galime naudoti pagreitinto gedimų laiko (AFT) regresinių modelių išskirtims ieškoti.

Raktiniai žodžiai : išskirtys, pagreitinto gedimų laiko (AFT) regresija, BP metodas, David Gather metodas, robastiniai įvertiniai.

Multiple Outliers Identification Method in Accelerated Failure Time (AFT) – Regression Models

Abstract

The purpose of the study is to give modifications of outliers identification methods for normal linear regression to more general case of accelerated failure time (AFT) regression and compare them. In this work three outlier search methods are investigated: BP and David Gather (DG) based on one of two parameter estimation methods: ordinary least squares and robust. The objective of the first part is to modify BP outlier search method for accelerated failure time (AFT) regression models. The objective of the second part is to compare all three methods using data generation and to provide practical examples. In many situations, BP outlier search method identifies outliers better then both DG methods (ordinary least squares and robust). Analysis of real data examples confirms simulation results. Obtained results proved, that BP outlier search method can be useful for outliers search in accelerated failure time (AFT) regression model.

Key words : outliers, accelerated failure time (AFT) regression, BP method, Davies Gather method, robust estimation.

Contents

1. Introduction

The main objective of the master thesis is to investigate a new outlier identification method in accelerated failure time (AFT) models and to create an outlier search procedure structure. The subject of outlier identification in accelerated failure time models is not widely investigated.

There are two different definitions of outliers. Firstly, the outlier region is defined as a set $out(X)$ such that with a very small probability at least one observation X_i from the sample is going to fall into $out(X)$ if the considered model holds. The first definition: an outlier is an observation which falls into the outlier region.

In the second case, the value x_i of X_i is an outlier if the probability distribution of X_i is different from the probability distribution given by the considered model - it is called *contaminants*.

If these two definitions are analyzed using *true* model, then in the first sense with a very small probability some outliers are possible, but using the second definition contaminants are absent. If contaminants are present, then the model does not hold for all observations. It is important to mention, that it is possible that contaminants do not fall into the outlier region. That means that contaminants are not necessarily outliers (in the first sense). Hence, the two notions are *different*. On the other hand, if the alternative distribution is concentrated in the outlier region, almost all outliers coincide using both definitions. In such cases, search methods for outliers and contaminants can be compared.

In this master thesis a new multiple outlier identification method for accelerated failure time (AFT) regression model based on robust estimation is presented. This method was developed based on the BP method [1] (described in the 3 section) and compared with the generalized Davies Gather method (which was applied to an accelerated failure time (AFT) regression model). Model's accuracy was established during data simulations for the AFT model, also it was tested on the real data and compared with the generalized Davies Gather method.

The rest of the work is organized as follows. Literature review is presented in Chapter 2. Outlier's definitions, outlier search in AFT models and the new model construction are presented in Chapter 3. Comparative analysis of the methods by simulation are given in

Chapter 4. The application of the new model to the real data is described in Chapter 5. Conclusions are presented in Chapter 6.

2. Literature review

For i.i.d. data $X_1, ..., X_n$ modelled by parametric models of the form $X_i \sim F(x;\theta)$, $\theta \in \Theta$, many outlier's identification methods were proposed in the statistical literature. However, the majority of the methods were given for the normal distribution (see [7, 12, 15, 33, 36, 39], see surveys in [4, 40]. Several methods were given for the exponential and gamma distributions, see [9, 18, 19, 21, 23–26, 41].

Many authors suppose that the number of possible outliers s are fixed. This assumption leads to only two possible conclusions: exactly s amount of observations are identified as outliers or none of the outliers are identified. It is more natural to consider methods which do not specify the number of suspected observations or at least specify the upper limit s. Such methods are not frequent and are modeled only for normal ([7, 17, 33]) or exponential samples ([20, 25, 26]). The only method which does not specify the upper limit s is the [12] method for normal samples. Bagdonavičius and Petkevičius proposed a method for outlier identification in samples from location-scale and shape-scale families of probability distributions [3].

Many outliers identification methods are proposed for the normal linear regression model. Outlier detection methods for linear regression model are described in the books of Fox [14], Barnett and Lewis [4], Riani and Atkinson [32], Rousseeuw and Leroy [22] and Chatterjee and Hadi [8].

It is common that estimated residuals based on ordinary least squares estimators (OLSE) of the regression parameter are used for many methods. If observations of internal or external studentized residuals are large, then corresponding observations are declared as outliers. Other methods consider more sophisticated functions of estimate residuals. Methods based on Cook's distance [10, 11], DFFITS [38], COVRATIO [5] statistics are implemented in standard statistical software. Peña [30] proposed to use well-chosen linear functions of Cook's distances. However, many authors remarked that if multiple outliers exist, then that all these OLSE based methods are not suitable – methods values have strong masking effect. Masking

effect means that not all outliers are identified as outliers ("masked"). Also some of these methods give large swamping effect, which means that many regular observations will be falsely identified as outliers.

Internal studentized residuals based on robust estimators of model parameters were studied in Rousseeuw and Van Zomeren work [34] and to reject outlier observations with large residuals values was proposed. The method described is much more convenient and accurate than methods based on ordinary least squares estimators (OLSE) in terms of masking and swamping effects.

The breakthrough in multiple outlier identification method improvement was made by Nurunnabi and Dai [29]. They modified Peñas method deleting “suspected” observations from the data and creating an obtained residuals set. Also Hadi, Rahmatullah Imon and Nurunnabi and Dai [16, 29, 31] studied a method based on group’s deleted residuals. If sets of “suspected” observations are selected properly, then latter methods gives good results. In the other hand, sometimes selecting proper sets leads to unexpected results, especially when datasets are large, i.e. using BACON [6] algorithm.

Bagdonavičius and Petkevičius [1] proposed a new multiple identification method on robust estimation, which is based on a result giving asymptotic properties of extreme studentized residuals.

In following sections generalizations of Davies-Gather (DG) method (given for the normal non-regression model) and Bagdonavičius-Petkevičius (BP) method (given for the normal regression model) to the case of accelerated failure time (AFT) models will be considered and their performance in terms of swamping and masking values will be compared.

3. Theoretic models and results

3.1. Accelerated failure time (AFT) model

Suppose that regression data are independent random vectors

$$(Y_1, (x^{(1)})^T), ..., (Y_n, (x^{(n)})^T), \tag{1}$$

where

$$x^{(i)} = (x_{i0}, x_{i1}, ..., x_{im})^T, \quad x_{i0} = 1, \tag{2}$$

is the vector of covariates for the ith object, and Y_i is the dependent variable.

Suppose that the AFT regression model

$$Y_i = \ln T_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i, \quad i = 1, ..., n,$$

is considered; here $\varepsilon_i/\sigma \sim F_0(x)$, F_0 is a specified function. Denote by f_0 the density of ε_i/σ.

The following three regression models are the most popular in survival analysis and reliability theory (see Table 1):

Distribution	$F_0(x)$
AFT-lognormal	$\Phi(x)$
AFT-Weibull	$1 - e^{-e^x}$
AFT-loglogistic	$\frac{1}{1+e^{-x}}$

Table 1: AFT-model distributions

Using matrix notation, the model is:

$$Y = X\beta + \varepsilon, \tag{3}$$

where $Y = (Y_1, ..., Y_n)^T$, $\beta = (\beta_0,\ \beta_1, ..., \beta_m)^T$, $\varepsilon = (\varepsilon_i, ..., \varepsilon_n)^T$, $X = [x_{ij}]$ is $n \times (m+1)$ matrix of covariates.

3.2. Definition of outliers and outlier regions for AFT models

Let us generalize the definition of the outlier region given in Davies and Gather [12] for non-regression normal data to the case of AFT models.

For an AFT model the right-sided α_n-outlier region can be defined as follows:

$$out_r(\alpha_n, F_0) = \{(y,x) \in (\mathscr{R} \times \mathscr{R}^{m+1} : y - \beta^T x > \sigma F_0^{-1}(1-\alpha_n)\}$$

and the left-sided α_n-outlier region is

$$out_l(\alpha_n, F_0) = \{(y,x) \in (\mathscr{R} \times \mathscr{R}^{m+1} : y - \beta^T x < \sigma F_0^{-1}(\alpha_n)\}.$$

The two-sided α-outlier region is

$$out(\alpha_n, F_0) =$$

$$\{(y,x) \in (\mathscr{R} \times \mathscr{R}^{m+1} : y - \beta^T x < \sigma F_0^{-1}(\alpha_n/2)\} \cup \{(y,x) \in (\mathscr{R} \times \mathscr{R}^{m+1} : y - \beta^T x > \sigma F_0^{-1}(1-\alpha_n/2)\}.$$

If f_0 is symmetric (as for AFT-lognormal and AFT-loglogistic models), then the two-sided outlier region is simpler:

$$out(\alpha_n, F_0) = \{(y,x) \in (\mathscr{R} \times \mathscr{R}^{m+1} : y - \beta^T x \in \mathscr{R}/[\sigma F_0^{-1}(\alpha/2)], \sigma F_0^{-1}(1-\alpha/2)]\}. \quad (4)$$

The α_n value is chosen supposing that if the AFT model (1) holds, then for a fixed $\alpha \in (0, 0.1]$

$$\mathbf{P}\Big\{ \bigcap_{i=1}^{n} \{(Y_i, x^{(i)}) \notin out_{\alpha_n}\} \Big\} = \big(\mathbf{P}\{(Y_i, x^{(i)}) \notin out_{\alpha_n}\}\big)^n = 1 - \alpha. \quad (5)$$

The equality (5) means that under the model (1) the probability that *none* of the random vectors $(Y_i, (x^{(i)})^T)$ falls into α_n - outlier region is $1 - \alpha$. This equality implies that

$$\alpha_n = 1 - (1-\alpha)^{1/n}. \quad (6)$$

The sequence α_n decreases from α to 0 as n goes from 1 to ∞.

The vector $(Y_i, (x^{(i)})^T)$ value is called *outlier* for a sample of size n if it falls into the outlier region out_{α_n}.

The number of outliers D_n under the model (1) has the binomial distribution $B(n, \alpha_n)$ and the expected number of outliers in the sample is $\mathbf{E}D_n = n\alpha_n$. Note that $\mathbf{E}D_n \to -\ln(1-\alpha) \approx \bar{\alpha}$ as $n \to \infty$. For example, if $\alpha = 0.05$ then $-\ln(1-\bar{\alpha}) \approx 0.05129$ and for $n \geq 10$ the expected number of outliers is approximately $0.051 \ll n$, i.e. it practically does not depend on n and is negligible with regard to the sample size n.

The definition of an outlier means that the value of $(Y_i, (x^{(i)})^T)$ is an outlier if it is far away from the regression plane (or line). The cause of an outlier may be unusual values of Y_i or $x^{(i)}$ (or both).

The value of $(Y_i, (x^{(i)})^T)$ is a high leverage point if the value of $x^{(i)}$ is far away from the bulk of the covariate values. Otherwise, it is a low leverage point ([34]). So a low leverage point is an outlier if Y_i takes an unusual value. A high leverage point is an outlier if it is far away from the regression plane, i.e. if it is a bad leverage point. A high leverage point is non-outlier if it is near to the regression plane, i.e. if it is a good leverage point.

So the outliers are bad high leverage points or low leverage points with unusual values of the dependent variable Y.

3.3. Generalization of Davies-Gather metod for AFT models

In section 3.1 was shown that by logarithmic transformation the AFT models are transformed to linear regression models with various distributions of error terms. The distribution of the error terms may be non-symmetric, cf. the AFT-Weibulll model.

After logarithmic transformation the data are

$$Y_i = \ln T_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \varepsilon_i, \quad i = 1, ..., n,$$

$\varepsilon_i/\sigma \sim F_0(x)$, F_0 is a specified function.

Let $\hat{\beta}$ and $\hat{\sigma}$ be robust estimators of the parameters β and σ. Denote by $\hat{Y}_i = \hat{\beta}^T x^{(i)}$, $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$, and $H = [h_{ij}]_{n\times n} = X(X^TX)^{-1}X^T$ the predicted values and unstandardized residuals, and leverage matrix, respectively. Set $h_i = h_{ii}$.

The studentized residuals are

$$r_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-h_i}}. \tag{7}$$

Define g_{n,α_n} using the condition

$$P\{r_i \le g_{n,\alpha}, i = 1, ..., n|H_0\} = P\{\max_{1\le i\le n} r_i \le g_{n,\alpha}|H_0\} = 1-\alpha, \tag{8}$$

So $g_{n,\alpha}$ is the α critical value of the random variable $\max_{1\le i\le n} r_i$.

Generalized Davies-Gather method for right outliers identification: right outliers are absent when $\max_{1\le i\le n} r_i \le g_{n,\alpha}$ and outliers exist when $\max_{1\le i\le n} r_i > g_{n,\alpha}$. They are selected finding the minimal i such that $r_{(i)} > g_{n,\alpha}$. Observations corresponding to $r_{(i)}, ..., r_{(n)}$ are declared as right outliers.

If estimators are equivariant, then the distribution of $\max_{1\le i\le n} r_i$ is parameter-free under the AFT model.

$h_{n,1-\alpha_n}$ is defined using the condition

$$P\{r_i \ge h_{n,1-\alpha}, i = 1, ..., n|H_0\} = P\{\min_{1\le i\le n} r_i \ge h_{n,1-\alpha}|H_0\} = 1-\alpha, \tag{9}$$

So $h_{n,1-\alpha}$ is the $1-\alpha$ critical value of the random variable $\min_{1\le i\le n} r_i$.

Generalized Davies-Gather method for left outliers identification: left outliers are absent when $\min_{1\le i\le n} r_i \ge h_{n,1-\alpha}$. The probability of such event is $1-\alpha$. Left outliers exist

when $\min_{1\le i\le n} r_i < h_{n,1-\alpha}$. They are selected finding the maximal i such that $r_{(i)} < h_{n,1-\alpha}$. Observations corresponding to $r_{(i)},...,r_{(n)}$ are declared as left outliers.

Let us consider two-sided case.

Suppose that the distribution of ε_i is symmetric with respect to 0.

Generalized Davies-Gather method for outliers identification (symmetric case): if $\max_{1\le i\le n}|r_i| \le g_{n,\alpha/2}$, then it is concluded that outliers are absent. If $\max_{1\le i\le n}|r_i| > g_{n,\alpha}$, then it is concluded that outliers exist. They are selected in the following way. Find the minimal i such that $|r|_{(i)} > g_{n,\alpha}$. Observations corresponding to $|r|_{(i)},...,|r|_{(n)}$ are declared as outliers. Outliers satisfying the inequality $r_{(i)} > g_{n,\alpha/2}$ are declared as right outliers and outliers satisfying the inequality $r_{(i)} < -g_{n,\alpha/2}$ are declared as left outliers.

Suppose that the distribution of ε_i is non-symmetric.

Generalized Davies-Gather method for left and right outliers identification (nonsymmetric distributions): if $\max_{1\le i\le n} r_i \le g_{n,\alpha/2}$ and $\min_{1\le i\le n} r_i \ge h_{n,1-\alpha/2}$, then it is concluded that the right outliers do not exist. The probability of such event is α. Otherwise, it is declared that outliers exist. Observations corresponding to i: $r_i > g_{n,\alpha/2}$ are declared as right outliers and observations corresponding to j: $r_j < h_{n,1-\alpha/2}$ are declared as left outliers.

3.4. Theoretic background for the BP identification method

Suppose that a c.d.f. F_0 belongs to the domain of attraction $\mathscr{G}_0$, i.e. normalizing constants $a_n > 0$ exist and $b_n \in \mathbf{R}$ such that $\lim_{n\to\infty} F_0^n(a_n x + b_n) = e^{-e^{-x}}$.

One of possible choices of the sequences $\{b_n\}$ and $\{a_n\}$ is

$$b_n = F_0^{-1}(1-\frac{1}{n}), \quad a_n = 1/(nf_0(b_n)). \tag{10}$$

Suppose that the function f_0 is not symmetric. The c.d.f. and p.d.f. are $1-F_0(-x)$ and $f_0(-x)$ respectively. Set

$$b_n^* = -F_0^{-1}(\frac{1}{n}), \quad a_n^* = 1/(nf_0(-b_n^*)). \tag{11}$$

In the particular case of the normal distribution equivalent form of $a_n = 1/b_n$ can be used. Expressions of b_n and a_n for some most used distributions are given in the following Table.

Distribution	$F_0(x)$	b_n	a_n
Normal	$\Phi(x) \in \mathscr{G}_0$	$\Phi^{-1}(1-1/n)$	$1/b_n$
Type I extreme value	$1-e^{-e^x} \in \mathscr{G}_0$	$\ln\ln n$	$e^{-b_n} = 1/\ln n$
Type II extreme value	$e^{-e^{-x}} \in \mathscr{G}_0$	$-ln(-ln(1-1/n))$	$e^{b_n}/(n-1)$
Logistic	$\frac{1}{1+e^{-x}} \in \mathscr{G}_0$	$\ln(n-1)$	$n/(n-1)$

Table 2: Classification table

Suppose that for any $\varepsilon > 0$

$$\lim_{x\to+\infty} x^{\varepsilon}[1-F_0(x)) = 0, \quad \lim_{x\to+\infty} x^{\varepsilon}[F_0(-x)) = 0. \tag{12}$$

Note that for all three considered probability distributions these conditions are satisfied:

1) Normal distribution: $F_0(x) = \Phi(x)$. Using the fact that $1-\Phi(x) \sim x^{-1}\varphi(x)$ as $x \to +\infty$, we have

$$\lim_{x\to+\infty} x^{\varepsilon}[\Phi(-x)] = \lim_{x\to+\infty} x^{\varepsilon}[1-\Phi(x)] = \lim_{x\to+\infty} x^{\varepsilon-1}\varphi(x) = 0.$$

2) Extreme value distribution: $F_0(x) = 1-e^{-e^x}$. We have

$$\lim_{x\to+\infty} x^{\varepsilon}[F_0(-x)] = \lim_{x\to+\infty} x^{\varepsilon}[1-e^{-e^{-x}}] = \lim_{x\to+\infty} x^{\varepsilon}[1-(1-e^{-x}+o(e^{-x}))] = \lim_{x\to+\infty} x^{\varepsilon}e^{-x} = 0.$$

$$\lim_{x\to+\infty} x^{\varepsilon}[1-F_0(x)] = \lim_{x\to+\infty} x^{\varepsilon}[e^{-e^x}] = \lim_{x\to+\infty} \frac{x^{\varepsilon}}{e^{e^x}} = 0.$$

3) Logistic distribution: $F_0(x) = \frac{1}{1+e^{-x}}$. We have

$$\lim_{x\to+\infty} x^{\varepsilon}[1-F_0(x)] = \lim_{x\to+\infty} x^{\varepsilon}[F_0(-x)] = \lim_{x\to+\infty} x^{\varepsilon}\frac{1}{1+e^x} = 0.$$

Note that $b_n = F_0^{-1}(1-\frac{1}{n}) \to \infty$ as $n \to \infty$ and if the conditions (12) are satisfied, then for any $\delta > 0$

$$\lim_{n\to\infty} \frac{b_n}{n^{\delta}} = 0, \quad \lim_{n\to\infty} \frac{a_n}{n^{\delta}} = 0. \tag{13}$$

Indeed, denoting $x = F_0^{-1}(1-\frac{1}{n})$, we have:

$$\lim_{n\to\infty} \frac{b_n}{n^{\delta}} = \lim_{n\to\infty} \frac{F_0^{-1}(1-\frac{1}{n})}{n^{\delta}} = \lim_{x\to+\infty} x[1-F_0(x)]^{\delta} =$$

$$\lim_{x\to+\infty} \left(x^{1/\delta}[1-F_0(x)]\right)^{\delta} = 0,$$

and applying l'Hopitals rule we have

$$0 = \lim_{n\to\infty} \frac{b_n}{n^{\delta}} = \lim_{n\to\infty} \frac{F_0^{-1}(1-\frac{1}{n})}{n^{\delta}} = \lim_{n\to+\infty} \frac{\frac{1}{n^2}}{\delta n^{\delta-1} f(F_0^{-1}(1-\frac{1}{n}))} =$$

$$\lim_{n\to+\infty}\frac{1}{\delta n^{\delta+1}f(b_n)}=\lim_{n\to+\infty}\frac{a_n}{\delta n^{\delta}}.$$

We supposed that $\varepsilon_i/\sigma \sim F_0(x)$. The c.d.f. of $|\varepsilon_i|/\sigma$ is $G_0(x)=F_0(x)-F_0(-x)$, $x\geq 0$ and the density is $g_0(x)=f_0(x)-f_0(-x)$.

If the conditions (12) are satisfied, then for any $\varepsilon>0$

$$\lim_{x\to+\infty}x^{\varepsilon}(1-G_0(x))=\lim_{x\to+\infty}x^{\varepsilon}(1-F_0(x)+F_0(-x))=0,\quad \lim_{x\to+\infty}x^{\varepsilon}G_0(-x)=0. \tag{14}$$

We used the fact that $F(-x)=0$ for all $x>0$.

So if $\tilde{b}_n=G_0^{-1}(1-1/n)$, $\tilde{a}_n=1/ng_0(\tilde{b}_n)$, then for any $\delta>0$

$$\lim_{n\to\infty}\frac{\tilde{b}_n}{n^{\delta}}=0,\quad \lim_{n\to\infty}\frac{\tilde{a}_n}{n^{\delta}}=0. \tag{15}$$

For symmetric distributions $G_0(x)=2F_0(x)-1$.

Let us suppose that the covariates are fixed or normally distributed for any fixed n. We need some conditions on the behaviour of the covariates, the c.d.f F_0 and estimators of the parameters β and σ as $n\to\infty$.

Denote by X_0 the sub-matrix of X without the first column of X and by $x_0^{(i)}$ the subvector of the vector $x^{(i)}$ without the first coordinate.

Conditions A. Consider a model that satisfies the following conditions:

1. $\frac{1}{n}X^TX\to Q$, where Q is non-degenerate symmetric matrix;
2. $\max_{1\leq i\leq n}||x^{(i)}||=O(b_n)$ as $n\to\infty$; here $||x^{(i)}||^2=\sum_{j=0}^{m}x_{ij}^2$;
3. $\hat{\beta}$ and $\hat{\sigma}$ are consistent estimators of β and σ, the limit distribution of $(\sqrt{n}(\hat{\beta}-\beta)^T,\sqrt{n}(\hat{\sigma}-\sigma))^T$ is non-degenerate.
4. $F_0\in\mathscr{G}_0$ and for any $\varepsilon>0$ $\lim_{x\to+\infty}x^{\varepsilon}[1-F_0(x))=0$, $\lim_{x\to+\infty}x^{\varepsilon}[F_0(-x))=0$, $\lim_{x\to\infty}x^3f_0(x)=0$, $\lim_{x\to\infty}xf_0(x)/(1-F_0(x))^{1/2}=0$.

Condition A1 is usual in asymptotic analysis of robust estimators of parameters. Condition A2 means that if covariates are unbounded, as $n\to\infty$, then they approach infinity not too quickly. If covariates are bounded, i.e. $\max_{1\leq i\leq n}||x^{(i)}||=O(1)$, then Condition A2 is automatically satisfied. Condition 4 is satisfied for all three considered distributions F_0.

Denote by $\hat{Y}_i = \hat{\beta}^T x^{(i)}$, $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$, and $H = [h_{ij}]_{n\times n} = X(X^TX)^{-1}X^T$ the predicted values, unstandardized residuals, and leverage matrix, respectively. Note that under Condition A

$$\max_{i,j} h_{ij} = \max_{i,j} \frac{(x^{(i)})^T}{\sqrt{n}} n(X^TX)^{-1} \frac{x^{(j)}}{\sqrt{n}} = O(\frac{b_n^2}{n}) \to 0 \tag{16}$$

as $n \to \infty$ because $n(X^TX)^{-1} \to Q^{-1}$ by Condition A 1, and $b_n/\sqrt{n} \to 0$ by (13). In particular,

$$r_0 = \min_i \sqrt{1-h_{ii}} \to 1 \tag{17}$$

Set $h_i = h_{ii}$. The studentized residuals are

$$r_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-h_i}}. \tag{18}$$

Order the random variables r_i:

$$r_{(1)} \le ... \le r_{(n)}.$$

Theorem 1. *Suppose that the c.d.f. F_0 belongs to the domain of attraction $\mathscr{G}_0$ and Conditions A hold. Then for fixed s*

$$\Big(\frac{r_{(n)} - b_n}{a_n}, \frac{r_{(n-1)} - b_n}{a_n}, ..., \frac{r_{(n-s+1)} - b_n}{a_n}\Big) \xrightarrow{d} L_0 \tag{19}$$

as $n \to \infty$; here

$$L_0 = (-\ln E_1, -\ln(E_1 + E_2), ..., -\ln(E_1 + ... + E_s)) \tag{20}$$

and $E_1, ..., E_s$ are i.i.d. standard exponential random variables.

Proof. The c.d.f. of the random variable ε_i/σ is $F_0 \in \mathscr{G}_0$. Set $\rho_i = \varepsilon_i/(\sigma\sqrt{1-h_i})$ and order these random variables:

$$\rho_{(1)} \le ... \le \rho_{(n)}.$$

For any $i = 1, ..., s$ the following equality holds:

$$\frac{r_{(n-i+1)} - b_n}{a_n} = \frac{r_{(n-i+1)} - \rho_{(n-i+1)}}{a_n} + \frac{\rho_{(n-i+1)} - b_n}{a_n}. \tag{21}$$

Note that $|r_i| \le |\rho_i| + |r_i - \rho_i|$, and let us consider the difference

$$r_i - \rho_i = \frac{1}{\sqrt{1-h_i}}\left(\frac{\hat{\varepsilon}_i}{\hat{\sigma}} - \frac{\varepsilon_i}{\sigma}\right) = \frac{1}{\sqrt{1-h_i}}\left(-\frac{(\hat{\beta}-\beta)^T x^{(i)}}{\hat{\sigma}} + \varepsilon_i(\frac{1}{\hat{\sigma}} - \frac{1}{\sigma})\right) =$$

$$\frac{1}{\hat{\sigma}\sqrt{1-h_i}}\left(-(\hat{\beta}-\beta)^T x^{(i)}+\frac{1}{\sqrt{n}}\frac{\varepsilon_i}{\sigma}\sqrt{n}(\hat{\sigma}-\sigma)\right). \tag{22}$$

By Condition A 3 and by (16)

$$\hat{\sigma}=\sigma+o_P(1), \quad \max_{1\le i\le n}\frac{1}{\sqrt{1-h_i}}\le\frac{1}{\min_{i,j}\sqrt{1-h_{ij}}}=(1-\max_{i,j}h_{ij})^{-1/2}=1+O(\frac{b_n^2}{n})=O(1). \tag{23}$$

By Condition A 3 $\sqrt{n}(\hat{\beta}-\beta)$ converges in distribution to a non-degenerate random variable and by Condition A 2 $\max_{1\le i\le n}||x^{(i)}||=O(b_n)$. Hence, applying the Cauchy-Schwartz inequality for any i we have

$$\max_{1\le i\le n}|(\hat{\beta}-\beta)^T x^{(i)}|\le||\sqrt{n}(\hat{\beta}-\beta)||\frac{1}{\sqrt{n}}\max_{1\le i\le n}||x^{(i)}||=O_P(\frac{b_n}{\sqrt{n}}). \tag{24}$$

Theorem 2.1.1 in [13] applied to the random variables $|\varepsilon_i|/\sigma$ implies that there exist a random variable V_1 with the c.d.f. $e^{-e^{-x}}$ such that

$$\frac{1}{\sigma\sqrt{n}}\max_{1\le i\le n}|\varepsilon_i|=(\tilde{b}_n+\tilde{a}_nV_1+o_P(\tilde{a}_n))/\sqrt{n}=o_P(1), \tag{25}$$

here $\tilde{b}_n=G_0^{-1}(1-1/n)$, $\tilde{a}_n=1/ng_0(b_n)$, $g_0=G_0'$. We used the result (15).

By Condition A 3 the sequence $\sqrt{n}(\hat{\sigma}-\sigma)$ converges in distribution to a non-degenerate random variable. Hence, the equality (22) and the results (23)-(25) imply that

$$\max_{1\le i\le n}|r_i-\rho_i|=O_P(\frac{b_n}{\sqrt{n}}).$$

The inequalities

$$\rho_i-\max_{1\le j\le n}|r_j-\rho_j|\le r_i\le\rho_i+\max_{1\le j\le n}|r_j-\rho_j|, \quad i=1,...,n,$$

imply

$$\rho_{(n-j+1)}-\max_{1\le i\le n}|r_i-\rho_i|\le r_{(n-j+1)}\le\rho_{(n-j+1)}+\max_{1\le i\le n}|r_i-\rho_i|, \quad j=1,...,s,$$

so

$$|r_{(n-j+1)}-\rho_{(n-j+1)}|\le\max_{1\le i\le n}|r_i-\rho_i|=O_P(\frac{b_n}{\sqrt{n}}).$$

and

$$\frac{r_{(n-i+1)}-\rho_{(n-i+1)}}{a_n}=O_P(\frac{b_n}{a_n\sqrt{n}}). \tag{26}$$

Condition A4 implies:

$$\lim_{n\to\infty}\frac{b_n}{\sqrt{n}a_n}=\lim_{x\to\infty}xf_0(x)/(1-F_0(x))^{1/2}=0.$$

Using (23) let us write the second term of the equality (21) in the form

$$\frac{\rho_{(n-i+1)}-b_n}{a_n}=\frac{\frac{\varepsilon_{(n-i+1)}}{\sigma(1-h_{(n-j+1)})^{1/2}}-b_n}{a_n}.\frac{\frac{\varepsilon_{(n-i+1)}}{\sigma}-b_n}{a_n}+O_P(\frac{b_n^2}{n})\frac{\varepsilon_{(n-i+1)}}{a_n\sigma} \tag{27}$$

Theorem 2.1.1 in [13] implies that

$$\frac{\varepsilon_{(n-i+1)}}{\sigma}=b_n+a_nV_j+o(a_n),$$

where V_j has non-degenerated distribution.

So

$$O_P(\frac{b_n^2}{n})\frac{\varepsilon_{(n-i+1)}}{a_n\sigma}=O(\frac{b_n^3}{na_n}).$$

Condition A4 implies:

$$\lim_{n\to\infty}\frac{b_n^3}{na_n}=\lim_{x\to\infty}x^3f_0(x)=0.$$

The equalities (21), (26), and (27) imply that for any $i=1,...,s$

$$\frac{|r|_{(n-i+1)}-b_n}{a_n}=\frac{\frac{|\varepsilon|_{(n-i+1)}}{\sigma}-b_n}{a_n}+o_P(1). \tag{28}$$

The s-dimensional random vector such that its ith component is the first term of the right side converges in distribution to L_0. It follows from Theorem 2.1.1 of [13] applied to the random variables $|\varepsilon_i|/\sigma$. Thus, the result (19) holds.

The proof is complete. □

Let us consider the case of random covariates. Note that in this case the random vectors $x_0^{(i)}$ are supposed to be i.i.d. with the mean $\mu=(\mu_1,...,\mu_m)^T=E(x_0^{(i)})$ and the matrix of the second moments $Q_0=(q_{jj'})=E(x_0^{(i)}(x_0^{(i)})^T)$, $j,j'=1,...,m$. Set $q_{00}=1$, $q_{0j}=q_{j0}=\mu_j$, $j=1,...,m$, and $Q=(q_{jj'})$, $j,j'=0,...,m$. The law of large numbers implies that $\frac{1}{n}X^TX\xrightarrow{P}Q$. So Condition A 1 is simply replaced by the following: Q is non-degenerate.

Condition A 2 is replaced by the following: $\max_{1\le i\le n}||x^{(i)}||=O_P(b_n)$ as $n\to\infty$. This condition is satisfied if the distribution of each $x_0^{(i)}$ has finite support which is natural in most practical situations. Even this is not necessary. Let us show that it is satisfied if the distribution of covariates is non-degenerate normal (it's support is not finite).

Proposition 1. *If* $x_0^{(i)}\sim N_m(\mu,\Sigma)$, $x_{ij}\sim N(\mu_j,\sigma_j^2)$, *then* $\max_{1\le i\le n}||x^{(i)}||=O(b_n)$ *as* $n\to\infty$.

Proof. Theorem 2.1.1 of [13] implies that for any $j = 1, ..., m$

$$\max_{1\le i\le n} |x_{ij}| = \mu_j + \sigma_j b_n + \frac{\sigma_j}{b_n} V_j + o_P(1),$$

where V_j is a random variable with the c.d.f. $e^{-e^{-x}}$. For any $M > 0$

$$P\Big\{\max_{1\le i\le n} |x_{ij}| \le M b_n\Big\} = P\Big\{V_j \le b_n \frac{b_n(M-\sigma_j) - \mu_j + o_P(1)}{\sigma_j}\Big\}.$$

For any fixed $M > \sigma_j$ the right side of the inequality converges to infinity in probability because $b_n \to +\infty$. Hence, for any $\varepsilon > 0$ there exist $M > 0$, $N = N(\varepsilon) > 0$ such that

$$P\{\max_{1\le i\le n} |x_{ij}| \le M b_n\} > 1 - \varepsilon \text{ as } n > N.$$

So for any $j = 1, ..., m$: $\max_{1\le i\le n} |x_{ij}| = O_P(b_n)$ as $n \to \infty$. It implies that

$$\max_{1\le i\le n} ||x^{(i)}|| = \max_{1\le i\le n} (\sum_{j=0}^{m} x_{ij}^2)^{1/2} \le \max_{1\le i\le n} \sum_{j=0}^{m} |x_{ij}| \le \sum_{j=0}^{m} \max_{1\le i\le n} |x_{ij}| = O_P(b_n). \qquad (29)$$

The proof is complete.

It can be shown that Theorem 1 holds in the case of random covariates, too.

3.5. Robust estimators for AFT regression models

The choice of the estimators $\hat{\beta}$ and $\hat{\sigma}$ is important when outlier detection problem is considered. When outliers exist the ML estimators from the complete sample are not stable.

Distribution	$K_0(x)$	d
Normal	$\Phi(x/\sqrt{2})$	2.2219
Type I extreme value	$1/(1+e^{-x})$	1.9576
Type II extreme value	$1/(1+e^{-x}$	1.9576
Logistic	$1 - \frac{(x-1)e^x+1}{(e^x-1)^2}$	1.3079

Table 3: Values of d for various probability distributions

For a fixed value of β, denote by $r^2_{(j)}(\beta)$ the ordered values of the residuals $\varepsilon_i^2 = (Y_i - \beta^T x^{(i)})^2$ (in increasing order). The LTS estimator minimizes the sum of squares

$$S_k(\beta) = \sum_{j=1}^{k} \varepsilon^2_{(j)}(\beta),$$

where so-called trimming constant k satisfies the condition $\frac{n}{2} < k \leq n$. The $n-k$ observations with the largest residuals will not affect the estimator.

The parameter σ is estimated by the statistic

$$\hat{\sigma} = Q_n = d\, W_{([0.25n(n-1)/2])}, \tag{30}$$

where $W_{ij} = |\hat{\varepsilon}_i - \hat{\varepsilon}_j|$, $1 \leq i < j \leq n$, $W_{(l)}$ is the lth order statistic from $C_n^2 = n(n-1)/2$ random variables W_{ij}.

The constant d has the form $d = 1/K_0^{-1}(5/8)$, where $K_0^{-1}(x)$ is the inverse of the c.d.f of $\varepsilon_1 - \varepsilon_2$, $\varepsilon_i \sim F_0$.

Expressions of $K_0^{-1}(x)$ and values d are given in Table 3.

If the distribution F_0 is symmetric around zero, then no further corrections are done.

If F_0 is not symmetric or the mean of ε_i is not equal to zero, then the estimator of β_0 is corrected. For example, in the case of the AFT-Weibull model the estimator of of β_0 has the form:

$$\hat{\beta}_0 = \tilde{\beta} + 0.33999 * \hat{\sigma},$$

where $\tilde{\beta}_0$ is estimated using the LTS method.

The robust LTS estimators were computed using `R` package `robustbase` ([27, 37]).

3.6. BP outlier search method

3.6.1. Identification of right outliers

Suppose that $F \in \mathscr{G}_0$. Let a_n, b_n be defined by (10). Set

$$U^+_{(n-i+1)}(n) = 1 - F_{\chi^2_{2i}}(2e^{-(r_{(n-i+1)}-b_n)/a_n}), \tag{31}$$

where $F_{\chi^2_{2i}}(x)$ is the c.d.f. of the chi-square distribution with $2i$ degrees of freedom. Set

$$U^+(n,s) = \max_{1\leq i\leq s} U^+_{(n-i+1)}. \tag{32}$$

If the AFT model with a specified F_0 and Conditions A hold, then Theorem 1 implies that the limit distribution (as $n \to \infty$) of the random variable $U^{(n)}(s)$ coincides with the distribution of the random variable

$$V^+(s) = \max_{1\leq i\leq s} V_i^+,$$

where $V_i^+ = 1 - F_{\chi^2_{2i}}(2(E_1 + ... + E_i))$, $i = 1, ..., s$, and $E_1, ..., E_s$ are i.i.d. standard exponential random variables. The random variables $V_1^+, ..., V_s^+$ are *dependent* identically distributed and the distribution of each V_i^+ is uniform: $V_i^+ \sim U(0,1)$.

Denote by $v_\alpha(s)$ the α critical value of the random variable $V^+(s)$. They are found many times simulating i.i.d. s standard exponential random variables and computing the values of $V^+(s)$.

Outlier search procedure begins with investigation of observations corresponding to the largest values of r_i. In Bagdonavičius and Petkevičius [1] article it is recommended to begin with five largest values. So take $s = 5$ and compute the value of the statistic $U^+(n,5) = \max_{1 \le i \le 5} U^+_{(n-i+1)}(n)$.

If $U^+(n,5) \le v^+_\alpha(5)$, then it is concluded that outliers are absent and no further investigation is done. If $U^+(n,5) > v^+_\alpha(5)$, then it is concluded that outliers exist and the following classification scheme is done.

Note that (see the classification scheme below) if $U^+(n,5) > v^+_\alpha(5)$, then minimum one observation is declared as an outlier. So the probability to declare absence of outliers does not depend on the following classification scheme.

Step 1. Set $d_1 = \max\{i \in \{1, ..., 5\} : U^+_{(n-i+1)}(n) > v^+_\alpha(5)\}$. If $d_1 < 5$, then classification is finished at this step: d_1 observations are declared as outliers, other observations are declared as non-outliers. If $d_1 = 5$, then it is possible that the number of outliers is higher than 5. Then the observation corresponding to $i = 1$ (i.e corresponding to $|r|_{(n)}$)) is declared as an outlier and we proceed to the step 2.

Step 2. The above written procedure is repeated taking $\max_{1 \le i \le 5} U^+_{(n-i)}(n-1) = U^+(n-1,5)$ instead of $U^+(n,5)$; here

$$U^+_{(n-i)}(n-1) = 1 - F_{\chi^2_{2i}}(2e^{-(r_{(n-i)} - b_{n-1})/a_{n-1}}), \quad i = 1, ..., 5,$$

Set $d_2 = \max\{i \in \{1, ..., 5\} : U^+(n-1)_{(n-i)} > v_\alpha(5)\}$. If $d_2 < 5$, the classification is finished and $d_2 + 1$ observations are declared as outliers.

If $d_2 = 5$, then it is possible that the number of outliers is higher than 6. In such case, the observation corresponding to the largest residual $|r|_{(n-1)}$ is declared as an outlier, in total 2 observations (i.e. corresponding to $r_{(n)}, r_{(n-1)}$) are declared as outliers at this step, but classification is not finished and the procedure should be repeated.

Classification finishes at the lth step, if $d_l < 5$. $(l-1)$ outliers were declaired in the previous steps and d_l outliers in the last one. The total number of observations declared as outliers is $l-1+d_l$. These observations correspond to $r_{(n)},...,r_{(n-d_l-l+2)}$.

From the computational point of view such method requires to find robust estimators and standardized residuals once. Linear time is needed for implementation of the classification procedure. Note that for fixed α ($\alpha = 0.05$, for example) only one critical value $v_\alpha(5)$ ($v_{0.05}(5) = 0.9853$, for example) is needed. Illustrative example showing simplicity of the BP method's application in 3.6.5 section.

3.6.2. Identification of left outliers

Let a_n^*, b_n^* be the normalizing constants defined by (11). Set

$$U_{(i)}^-(n) = 1 - F_{\chi^2_{2i}}(2e^{(r_{(i)}-b_n^*)/a_n^*}), \tag{33}$$

where $F_{\chi^2_{2i}}(x)$ is the c.d.f. of the chi-square distribution with $2i$ degrees of freedom. Set

$$U^-(n,s) = \max_{1\le i\le s} U_{(i)}^-(n). \tag{34}$$

If the AFT model with a specified F_0 and Conditions A hold, then Theorem 1 implies that the limit distribution (as $n\to\infty$) of the random variable $U^-(n,s)$ coincides with the distribution of the random variable

$$V^-(s) = \max_{1\le i\le s} V_i^-,$$

where $V_i^- = 1 - F_{\chi^2_{2i}}(2(E_1+...+E_i))$, $i = 1,...,s$, and $E_1,...,E_s$ are i.i.d. standard exponential random variables. The random variables $V_1^-,...,V_s^-$ are *dependent* identically distributed and the distribution of each V_i^- is uniform: $V_i^- \sim U(0,1)$.

Denote by $v_\alpha^-(s)$ the α critical value of the random variable $V^-(s)$. They are found simulating i.i.d. s standard exponential random variables and computing the values of $V^-(s)$. The critical values $v_\alpha^-(s)$ are approximated by the critical values $v_\alpha^-(s) = v_\alpha^+(s)$.

The left outliers search method coincides with the right outliers search method replacing + to - in all formulas.

3.6.3. Identification of two-sided outliers for symmetric distributions

Let a_n, b_n by defined by (10). If $F \in \mathscr{G}_0$, then set

$$U_{(n-i+1)}(n) = 1 - F_{\chi^2_{2i}}(2e^{-(|r|_{(n-i+1)} - b_{2n})/a_{2n}}). \tag{35}$$

Set

$$U(n,s) = \max_{1 \le i \le s} U_{(n-i+1)}(n). \tag{36}$$

Denote by $v_\alpha(n,s)$ the α critical value of the statistic $U(n,s)$. Theorem 1 and Remark 2 ([3]) imply that the limit distribution (as $n \to \infty$) of the random variable $U(n,s)$ coincides with the distribution of the random variable $V^+(s)$. The critical valuses $v_\alpha(n,s)$ are approximated by the critical values $v_\alpha(s) = v^+_\alpha(s)$. The outliers search method coincides with the right outliers search method skipping upper index + in all formulas.

3.6.4. Identification of two-sided outliers for non-symmetric distributions

Suppose that the function f_0 is not symmetric (for example AFT-Weibull). Let a_n, b_n, a^*_n, b^*_n be defined by 10, 11.

Begin outlier search using observations corresponding to the largest and the smallest values of $\hat{r}_i$. Compute the values of statistics $U^-(n,s)$ and $U^+(n,s)$. If $U^-(n,s) \le v_{\alpha/2}(s)$ and $U^+(n,s) \le v_{\alpha/2}(s)$, then it is concluded that outliers are absent and no further investigation is done.

If $U^-(n,s) > v_{\alpha/2}(s)$ or $U^+(n,s) > v_{\alpha/2}(s)$, then it is concluded that outliers exist. If $U^-(n,s) > v_{\alpha/2}(s)$, then left outliers are searched as in Section 3.6.2. If $U^+(n,s) > v_{\alpha/2}(s)$, then right outliers are searched in Section 3.6.1. The difference is that α is replaced by $\alpha/2$ in all formulas.

3.6.5. Illustrative right outliers example

To illustrate simplicity of the BP method, let us consider an illustrative example of its application (sample size $n = 30$, $r = 7$ outliers). The sample size of $n = 30$ from Weibull distributuion was generated. The last 7 observations were replaced by outliers. The regression data (x_j, y_j) and the robust residuals $r_{(n-j+1)}$ are presented below:

j	X_j	Y_j	$r_{(n-j+1)}$	j	X_j	Y_j	$r_{(n-j+1)}$
1	4.3913260	5.47390926	0.23939917	16	1.6317853	0.83141322	-1.46632251
2	0.6889999	1.32292754	-0.17412374	17	-1.1770141	-1.80546474	-1.31302904
3	0.3228709	0.94531960	-0.18463099	18	3.7063257	3.11150483	-1.30474626
4	-4.1343077	-3.62693403	-0.30189729	19	3.5882593	4.00311988	-0.37657768
5	-2.9045486	-3.54841355	-1.34765425	20	-1.4872407	-0.36049057	0.26802210
6	0.7551341	2.41317329	0.74563577	21	-0.8830237	-0.03498023	0.01697442
7	-3.2840811	-3.16817867	-0.65679695	22	-1.7204924	-1.59618227	-0.63758299
8	1.1929285	1.77817425	-0.21775959	23	4.8576262	5.66352051	-0.01722873
9	-2.9889060	-2.50644893	-0.31878905	24	-2.8707830	6.87729607	8.14548145
10	0.3017340	-0.02289087	-1.03470683	25	0.1237114	6.88755257	5.32744173
11	2.8610408	2.49645136	-1.08297317	26	-4.4949313	6.85646198	9.84835468
12	-5.3709688	-4.17428219	0.34681885	27	1.9828775	6.74523635	3.55357514
13	0.9351036	1.49180254	-0.24342270	28	0.8563729	6.68461173	4.49191429
14	7.6619640	7.39084309	-1.10107093	29	4.1665989	6.80239476	1.67502211
15	-2.5825133	-1.56844291	0.16704422	30	-0.4780308	6.90775528	5.88999271

Steps of the classification procedure by the BP method is presented in the table below (see Table 4)

$U_{(30)}(30)$	$U_{(29)}(30)$	$U_{(28)}(30)$	$U_{(27)}(30)$	$U_{(26)}(30)$	$U(30,5)$
1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
$U_{(29)}(29)$	$U_{(28)}(29)$	$U_{(27)}(29)$	$U_{(26)}(29)$	$U_{(25)}(29)$	$U(29,5)$
1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000
$U_{(28)}(28)$	$U_{(27)}(28)$	$U_{(26)}(28)$	$U_{(25)}(28)$	$U_{(24)}(28)$	$U(28,5)$
0.9999999	1.0000000	1.0000000	1.0000000	0.9999967	1.0000000
$U_{(27)}(27)$	$U_{(26)}(27)$	$U_{(25)}(27)$	$U_{(24)}(27)$	$U_{(23)}(27)$	$U(27,5)$
0.9999991	1.0000000	1.0000000	0.9999240	0.4246993	1.0000000

Table 4: Classification table

First, we compute (see line 1 of Table: 4) of the statistic $U(30,5) = \max_{1\leq i\leq 5} U^{-}_{(30-i+1)}(30) = 1$. So, $U(20,5) = 1 > 0.9853 = v_{0.05}(5)$, null hypothesis was rejected and that concludes, that outliers exist. The outliers search can be started.

Step 1. The inequality $U_{(26)}(30) = 1.0000 > 0.9583 = v_{0.005}(5)$ so we conclude that outliers exist. The number d_1 is equal to the selected limit 5. So it is possible that the number

of outliers might be greater than 5. We reject the largest in absolute value 30th observation as an outlier and continue the search of outliers.

Step 2. The inequality $U_{(25)}(29) = 1.0000 > 0.9583 = v_{0.005}(5)$ implies that $d_2 = 5$. So we declare as an outlier the observation with the second largest value of the residual. So two observations are declared as outliers. We continue the search of outliers.

Step 3. The inequality $U_{(24)}(28) = 0.9999967 > 0.9583 = v_{0.005}(5)$ implies that $d_3 = 5$. The third largest in absolute value observation is declaired as an outlier. The search of outliers continues.

Step 4. The inequality $U_{(23)}(27) = 0.4246993 < 0.9583 = v_{0.005}(5)$ and $U_{(24)}(27) = 0.9999240 > 0.9583 = v_{0.005}(5)$ imply that $d_4 = 4$. So four additional observations are declared as outliers. The outlier search is finished. In all, seven observations were declared as outliers, as we expected.

4. Practical investigation

4.1. Simulation scheme

The simuation scheme was used to compare outlier identification methods. It is described as follows.

Suppose that n is the size of a sample. If we want the first r observations to be outliers with the probability equals one, we must generate observations using any probability distribution concentrated with the probability equals one in the outlier region. We shall call such observations contaminated outliers (c-outliers). The remaining $n-r$ observations are generated using the AFT regression model (3). Realizations of these $n-r$ observations may be outliers with very small probability.

For fixed $n = 20, 50, 100, 1000$ we generated univariate i.i.d. $z_i \sim N(0,1), i = 1, ..., n$, $W_i \sim Weibull(\nu, \theta_i)$ ($LL_i \sim Loglogistic(\nu, \theta_i), LN_i \sim Lognormal(\mu_i, \sigma)$) where shape $\nu = 1.5$ and scale $\theta_i = e^{1+z_i}$ (for lognormal $\mu_i = 1 + z_i, \sigma = 1$). For $i = r+1, ..., n$ we generated $\varepsilon_i \sim F_0(x)$ (see Table:1 F_0 values for different cases). For $i = 1, ..., r$ we generated i.i.d. random variables ε_i following the truncated exponential distribution $\mathscr{E}(\theta, z_{\alpha_n})$ with various values of the scale parameter θ and left truncation values z_{α_n} (for example in AFT-Weibull

case $z_{\alpha_n} = log(-log(\alpha_n))$, AFT-Loglogistic $z_{\alpha_n} = log(1-\alpha_n) - log(\alpha_n)$, AFT-Lognormal $z_{\alpha_n} = \Phi^{-1}(1-\alpha_n)$, where $\Phi^{-1}(x)$ is the inverse distribution function of the standard normal distribution), so ε_i fall with the probability one into the region $(z_{\alpha_n}, +\infty)$. Then we computed the values of $Y_i = W_i + \varepsilon_i$ (i.e. we took $\beta_o = 0, \beta_1 = 1$). Note that the inequality $|\varepsilon_i| > z_{\alpha_n}$ is equivalent to the inequality $|Y_i - W_i| > z_{\alpha_n}$, so for $i = 1, ..., r$ the observations (Y_i, W_i) fall into the α_n-outlier region. If θ increases, then the mean distance from the border of the outlier region and the spread of outliers in this region increase.

The $\alpha = 0.05$ was fixed for the definition of outlier region (see (5) formula).

The simulations presented below were performed using *R* software. For any fixed sample size *n*, fixed number of c-outliers *r*, and fixed alternative or null hypothesis, the performance of each method was investigated using 10000 simulated samples. The robust LTS estimators were computed using `R` package `robustbase` ([27, 37]).

4.2. AFT-Weibull regression model

4.2.1. Comparison between BP and Davies Gather for right outlier search method

Outlier identification methods are ideal if each outlier can be detected and each non-outlier can be declared as a non-outlier. In practice it is impossible to detect outliers with the probability one. Two errors are possible: (a) an outlier is not declared as such (masking effect); (b) a nonoutlier is declared as an outlier (swamping effect). In this work name "masking" is for the mean number of non-detected c-outliers and name "swamping" is for the mean number of "normal" observations declared as outliers in the simulated samples. If c-outliers are absent, then swamping is simply the mean number of observations declared as outliers.

Method	n=20	n=50	n=100	n=1000
	α	α	α	α
BP	0.0488	0.0487	0.0483	0.0554
DG rob	0.049	0.0467	0.049	0.0539
DG	0.0482	0.0515	0.0479	0.0511

Table 5: The proportions $\alpha = p_O$ values given that c-outliers are absent for Right AFT-Weibull regression model

Table 5 shows the values of the proportions p_O of samples declared as having outliers.

From the Table 5 we can see that when n increase p_O values decreases. For all three methods (BP, DG robust, DG) all α values are around 0.05, this shows us that when we do not have c-outliers in small number samples we can find some outliers.

Let's compare these three methods with some outliers (r) and check how these methods work fixing different sample sizes.

r	2							
Method/ θ	0.05		0.01		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.942 (out of 2)	0.0949 (out of 18)	0.832 (out of 2)	0.0977 (out of 18)	0.3268 (out of 2)	0.0935 (out of 18)	0.0714 (out of 2)	0.1042 (out of 18)
DG rob	1.362 (out of 2)	0.0123 (out of 18)	1.2415 (out of 2)	0.0163 (out of 18)	0.5083 (out of 2)	0.0158 (out of 18)	0.1336 (out of 2)	0.0166 (out of 18)
DG	1.4947 (out of 2)	0 (out of 18)	1.3611 (out of 2)	0 (out of 18)	0.6654 (out of 2)	0.0001 (out of 18)	0.3194 (out of 2)	0.0001 (out of 18)

Table 6: The masking and swamping values for Right AFT-Weibull model $(n = 20)$

r	5							
Method/ θ	0.05		0.01		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.2684 (out of 5)	0.1333 (out of 45)	0.2785 (out of 5)	0.1278 (out of 45)	0.0168 (out of 5)	0.1625 (out of 45)	0.0018 (out of 5)	0.1648 (out of 45)
DG rob	2.2957 (out of 5)	0.0054 (out of 45)	2.3519 (out of 5)	0.0039 (out of 45)	0.4148 (out of 5)	0.005 (out of 45)	0.0839 (out of 5)	0.0068 (out of 45)
DG	4.0742 (out of 5)	0.0001 (out of 45)	4.0593 (out of 5)	0 (out of 45)	1.4319 (out of 5)	0 (out of 45)	1.0102 (out of 5)	0 (out of 45)

Table 7: The masking and swamping values for Right AFT-Weibull model $(n = 50)$

From the results we can see that masking values decrease if the parameters characterizing remoteness of c-outliers increase. For $n = 20, 50, 100,$ and 1000 the masking values are presented accordingly in Tables 6, 7, 8 and 9.

r	10							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.2661 (out of 10)	0.1038 (out of 90)	0.2519 (out of 10)	0.1081 (out of 90)	0.2576 (out of 10)	0.1064 (out of 90)	0.0028 (out of 10)	0.1134 (out of 90)
DG rob	2.4844 (out of 10)	0.0028 (out of 90)	2.4572 (out of 10)	0.0018 (out of 90)	2.5641 (out of 10)	0.0018 (out of 90)	0.1962 (out of 10)	0.0021 (out of 90)
DG	8.8944 (out of 10)	0 (out of 90)	8.9111 (out of 10)	0 (out of 90)	8.8874 (out of 10)	0 (out of 90)	2.7243 (out of 10)	0 (out of 90)

Table 8: The masking and swamping values for Right AFT-Weibull model ($n = 100$)

r	100							
Method/ θ	0.001		0.005		0.02		0.05	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0 (out of 100)	0.0196 (out of 900)	0 (out of 100)	0.022 (out of 900)	0 (out of 100)	0.0206 (out of 900)	0 (out of 100)	0.0199 (out of 900)
DG rob	0 (out of 100)	0.0001 (out of 900)	0 (out of 100)	0.0001 (out of 900)	0.0001 (out of 100)	0.0001 (out of 900)	0.0001 (out of 100)	0 (out of 900)
DG	99.8376 (out of 100)	0 (out of 900)	99.8391 (out of 100)	0 (out of 900)	99.8401 (out of 100)	0 (out of 900)	99.8215 (out of 100)	0 (out of 900)

Table 9: The masking and swamping values for Right AFT-Weibull model ($n = 1000$)

Let's describe the results for small ($n = 20$), medium ($n = 100$), and large (n = 1000) samples separately.

1) $n = 20$. In Table 6 data shows that comparing masking values (in terms of the numbers of nonidentified c-outliers) of BP and DG robust methods the latter method's values are larger than BP method's, for all θ (when many outliers are concentrated near the outlier region border). For all small $\theta \leq 0.1$ more than half of outliers were found using BP method, but only about third of outliers were found using DG robust method. However increasing θ value (more than 0.1) both methods show better results. Talking about DG method without robust estimation masking values with all θ values are much bigger then BP or DG robust methods. DG method without robust estimation shows worst results then DG with robust estimation. Therefore, when $n = 20$ and θ are small BP method shows better results at identifying outliers, but increasing θ both BP and DG robust methods begin to show good

results, DG method without robust estimation identifies outliers badly. It is important to notice that for all θ swamping values of DG methods are smaller than BP method's, however for both methods swamping values are small enough to conclude that BP method shows better results, because masking values for all θ are significantly smaller than DG and DG robust methods.

2) $n = 100$. Table 8 data shows that comparing masking values of BP and DG robust methods the latter method's values are larger than BP method's, for all θ. For all small $\theta \leq 0.1$ DG method masking values are around 2.5, which means that less than seven and a half outliers out of ten were identified, however masking values of DG robust method decreases to approximately 0.2 when θ value increases from 0.1 to 1. Using BP method more than 97% of outliers were identified, for all θ. Talking about DG method without robust estimation masking values are very high near outlier border, when $\theta = 1$ masking value are also big and equal 2.7243. Therefore, when $n = 100$ and θ are small BP method shows better results at identifying outliers, but increasing θ BP and DG robust methods begin to show good results. It is important to notice that for all θ swamping values of DG methods are smaller than BP method's, DG without robust estimation method swamping values with all θ are equal 0. However all three methods swamping values are small enough to conclude, that BP method shows better results than DG or DG robust methods.

3) $n = 1000$. In Table 9 data for all θ shows that all masking values of BP method are equal 0, however DG robust method majority of masking values are around 0. This shows, that for all θ BP method finds all outliers and DG robust method finds almost all outliers. DG method without robust estimation shows very bad results, masking values with all θ values are very big (near 100). For all θ on all methods swamping values are around 0, but it is important to notice that DG methods swamping values are smaller than BP method. Therefore, BP and DG robust methods swamping values are particularly small and with large sample size (n) detects outliers similar. DG method without robust estimation can not detect correctly outliers.

Conclusion: In most considered situations, the BP method is the best outlier identification method. However, with large simple size $n = 1000$ DG robust method have similar performance.

4.2.2. Comparison between BP and Davies Gather for left outlier search method

For the left outlier search in AFT-Weibull regression model we can see (see Table 10), that for all methods (BP, DG robust and DG) with different n, p_O values are near 0.05 (theoretical α). Table 10 gives the values of the proportions p_O of samples declared as having outliers.

Method	n=20	n=50	n=100	n=1000
	α	α	α	α
BP	0.0396	0.0345	0.041	0.0572
DG rob	0.0483	0.0555	0.0488	0.0539
DG	0.0507	0.0518	0.0486	0.0521

Table 10: The proportions $\alpha = p_O$ values given that c-outliers are absent for Left AFT-Weibull regression model

Let's compare these three methods with some outliers (r) and check how this methods works with different sample sizes n and different truncated exponential distribution scale parameter θ. (see Tables: 11, 12, 13 and 14)

r	2							
Method/ θ	0.05		0.1		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.1345 (out of 2)	0.2583 (out of 18)	0.1214 (out of 2)	0.2722 (out of 18)	0.0619 (out of 2)	0.2512 (out of 18)	0.0158 (out of 2)	0.2569 (out of 18)
DG rob	0.6435 (out of 2)	0.0165 (out of 18)	0.6222 (out of 2)	0.019 (out of 18)	0.321 (out of 2)	0.0194 (out of 18)	0.1057 (out of 2)	0.0176 (out of 18)
DG	1.2297 (out of 2)	0 (out of 18)	1.2417 (out of 2)	0 (out of 18)	1.0436 (out of 2)	0 (out of 18)	0.8712 (out of 2)	0 (out of 18)

Table 11: The masking and swamping values for Left AFT-Weibull model $(n = 20)$

Masking decreases if the parameters characterizing remoteness of c-outliers increase. Therefore describe the results for small ($n = 20$), medium ($n = 100$), and large ($n = 1000$) samples separately.

1) $n = 20$. In Table 11 data shows that comparing masking values (in terms of the numbers of nonidentified c-outliers) of BP and DG robust methods the latter method's values are

r	5							
Method/ θ	0.05		0.1		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.0026 (out of 5)	0.29 (out of 45)	0.0029 (out of 5)	0.2867 (out of 45)	0.0023 (out of 5)	0.2857 (out of 45)	0.0007 (out of 5)	0.2841 (out of 45)
DG rob	0.4442 (out of 5)	0.0123 (out of 45)	0.4343 (out of 5)	0.0142 (out of 45)	0.1372 (out of 5)	0.0125 (out of 45)	0.0339 (out of 5)	0.0113 (out of 45)
DG	4.3436 (out of 5)	0 (out of 45)	4.3928 (out of 5)	0 (out of 45)	3.4885 (out of 5)	0 (out of 45)	3.278 (out of 5)	0 (out of 45)

Table 12: The masking and swamping values for Left AFT-Weibull model $(n = 50)$

r	10							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.0013 (out of 10)	0.2734 (out of 90)	0.0013 (out of 10)	0.271 (out of 90)	0.001 (out of 10)	0.2624 (out of 90)	0.001 (out of 10)	0.2584 (out of 90)
DG rob	0.2036 (out of 10)	0.009 (out of 90)	0.2154 (out of 10)	0.0099 (out of 90)	0.2125 (out of 10)	0.0087 (out of 90)	0.0436 (out of 10)	0.0094 (out of 90)
DG	9.9703 (out of 10)	0 (out of 90)	9.9794 (out of 10)	0 (out of 90)	9.9827 (out of 10)	0 (out of 90)	8.4776 (out of 10)	0 (out of 90)

Table 13: The masking and swamping values for Left AFT-Weibull model $(n = 100)$

r	100							
Method/ θ	0.001		0.005		0.02		0.05	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0 (out of 100)	0.2033 (out of 900)	0 (out of 100)	0.2017 (out of 900)	0 (out of 100)	0.1978 (out of 900)	0 (out of 100)	0.1964 (out of 900)
DG rob	0 (out of 100)	0.0052 (out of 900)	0 (out of 100)	0.0063 (out of 900)	0 (out of 100)	0.006 (out of 900)	0 (out of 100)	0.0062 (out of 900)
DG	100 (out of 100)	0 (out of 900)	100 (out of 100)	0 (out of 900)	100 (out of 100)	0 (out of 900)	100 (out of 100)	0 (out of 900)

Table 14: The masking and swamping values for Left AFT-Weibull model $(n = 1000)$

larger than BP method's, for all θ (when many outliers are concentrated near the outlier region border). DG method without robust estimation show worst results than DG method

with robust estimation or BP method. In other hand, it is important to notice that for all θ swamping values of DG methods are smaller than BP method's. However, BP method has larger swamping effect and less masking effect than DG robust or DG without robust estimation methods and that mean BP method is better than DG methods.

2) $n = 100$. Table 13 data shows that comparing masking values of BP and DG robust methods the latter method's values are larger than BP method's, for all θ. For all small θ ($\theta \leq 0.1$). DG robust method masking values are around 0.2, θ increasing to 1 DG robust masking value decreases around zero. BP method masking values are around 0 for all θ. DG without robust estimation method masking values are very high (around 10), when θ increase masking values decrease very slowly. Therefore, when $n = 100$ and θ are small BP method shows better results at identifying outliers, but increasing θ both BP and DG robust methods begin to show good results. DG wihout robust estimation do not show good results and this method do not correctly detect outliers for medium sample sizes. It is important to notice that for all θ swamping values of DG methods are smaller than BP method's.

3) $n = 1000$. In Table 14 data for all θ shows that all masking values of BP and DG robust methods are equal 0. This shows, that for all θ both methods finds all outliers. Different view shows DG method without robust estimation, whit all θ values masking value are equal 100. That means DG method without robust estimation can not find any outlier. In other hand, for all θ swamping values on DG methods are around 0, but BP method is around 0.20. Although, BP and DG robust methods swamping values are particularly small (around 0) and both methods are simillar.

Conclusion: Therefore, the results are very similar as Right AFT-Weibull event (see Section: 4.2.1). In most considered situations, the BP method is the best outlier identification method. It is important to notice that DG robust method have the same performance with large simple size $n = 1000$ on left outlier search method. DG method without robust estimation have big masking values with different sample sizes and this method can not use for searching outliers.

4.2.3. Comparison between BP and Davies Gather for two-sided outlier search method

In this section was investigated medium size value ($n = 100$) outlier search method. This outlier search method was created by Section 3.6.4, so for detailed analysis see: 4.2.1 and

4.2.2 Sections results.

Let's compare BP and Davies-Gather methods with 10 outliers ($r_{left} = 5, r_{right} = 5$) and check how this three methods works for two-sided method with different truncated exponential distribution scale parameter θ. (see Tables: 15 and 16).

r_{right}	5							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.0133 (out of 5)	0.121 (out of 95)	0.0158 (out of 5)	0.1165 (out of 95)	0.0178 (out of 5)	0.1108 (out of 95)	0.0094 (out of 5)	0.1171 (out of 95)
DG rob	0.5549 (out of 5)	0.0026 (out of 95)	0.5573 (out of 5)	0.0017 (out of 95)	0.5771 (out of 5)	0.0021 (out of 95)	0.5643 (out of 5)	0.0024 (out of 95)
DG	4.8788 (out of 5)	0 (out of 95)	4.8757 (out of 5)	0 (out of 95)	4.8769 (out of 5)	0 (out of 95)	4.8848 (out of 5)	0 (out of 95)

Table 15: The masking and swamping values for right outlier search in AFT-Weibull model ($n = 100$)

r_{left}	5							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.0001 (out of 5)	0.2655 (out of 95)	0.0002 (out of 5)	0.2538 (out of 95)	0.0001 (out of 5)	0.2562 (out of 95)	0.0001 (out of 5)	0.2518 (out of 95)
DG rob	0.0331 (out of 5)	0.0104 (out of 95)	0.0358 (out of 5)	0.0093 (out of 95)	0.0369 (out of 5)	0.0091 (out of 95)	0.0352 (out of 5)	0.0109 (out of 95)
DG	0.7906 (out of 5)	0.0005 (out of 95)	0.7455 (out of 5)	0.0004 (out of 95)	0.7628 (out of 5)	0.0007 (out of 95)	0.7618 (out of 5)	0.0008 (out of 95)

Table 16: The masking and swamping values for left outlier search in AFT-Weibull model ($n = 100$)

Tables 15, 16 shows that comparing masking values of DG robust and BP methods the latter method's values are larger than BP method's, for all θ (for left and right outliers). For left outlier search BP method masking value are near 0 with all θ values while DG robust method masking values are approximately 0.35. DG method without robust estimation have much biggest masking values than DG robust or BP methods. Talking about right outliers search BP method masking values increasing from 0.0133 to 0.0178 when θ increase from

0.01 to 0.1, but when $\theta = 1$ masking value decreasing to 0.0094. For DG robust method masking values are bigger than BP method more than 50 times, DG method without robust estimation in this case works much worst and masking values are around 5 with all θ values. It is important to notice that for all θ (for left and right outlier search) swamping values of DG methods are smaller than BP method's. Therefore all three models swamping values are small enough to conclude, that BP method show better results than DG robust and DG methods.

Conclusion: This two-sided outlier search method showed very similar results as previously investigated right and left outlier search methods (see 4.2.1 and 4.2.2 Sections results). With medium ($n = 100$) sample size BP method are better than DG and DG robust methods, especially when outliers are near outlier region border ($\theta < 1$).

4.3. AFT-Loglogistic regression model

4.3.1. Comparison between BP and Davies Gather for right outlier search method

In this section performance of BP and Davies Gather methods for AFT Loglogistic regression model was analyzed. Table 17 shows the values of the proportionsp_O of samples declared as having outliers. All three methods α values are near 0.05, which means that swamping effect should be minimal. Analyzing BP methods value, when $n = 1000$, we notice that shown value $\alpha = 0.0787$, this means that p_O is bigger than theoretical α.

Method	n=20	n=50	n=100	n=1000
	α	α	α	α
BP	0.0398	0.0417	0.0526	0.0787
DG rob	0.0515	0.0522	0.0507	0.0488
DG	0.0525	0.0523	0.0527	0.0498

Table 17: The proportions $\alpha = p_O$ given that c-outliers are absent for Right AFT-Loglogistic regression model

Let's check how BP, DG robust and DG methods works with some outliers r and investigate swamping and masking effect. For $n = 20, 50, 100$ and 1000 the masking values are presented in Tables: 18, 19, 20, 21.

r	2							
Method/ θ	0.05		0.1		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	1.1484 (out of 2)	0.1818 (out of 18)	1.1637 (out of 2)	0.1829 (out of 18)	0.8503 (out of 2)	0.1987 (out of 18)	0.3336 (out of 2)	0.2134 (out of 18)
DG rob	1.7433 (out of 2)	0.0079 (out of 18)	1.7467 (out of 2)	0.0099 (out of 18)	1.4259 (out of 2)	0.0124 (out of 18)	0.6716 (out of 2)	0.0164 (out of 18)
DG	1.9021 (out of 2)	0 (out of 18)	1.9107 (out of 2)	0 (out of 18)	1.7627 (out of 2)	0 (out of 18)	1.1635 (out of 2)	0 (out of 18)

Table 18: The masking and swamping values for Right AFT-Loglogistic model $(n = 20)$

r	5							
Method/ θ	0.05		0.1		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.1995 (out of 5)	0.2952 (out of 45)	0.2028 (out of 5)	0.2824 (out of 45)	0.1018 (out of 5)	0.2852 (out of 45)	0.0379 (out of 5)	0.2988 (out of 45)
DG rob	4.6263 (out of 5)	0.002 (out of 45)	4.6463 (out of 5)	0.0028 (out of 45)	3.5691 (out of 5)	0.0027 (out of 45)	1.4312 (out of 5)	0.0057 (out of 45)
DG	4.9857 (out of 5)	0 (out of 45)	4.9829 (out of 5)	0 (out of 45)	4.8223 (out of 5)	0 (out of 45)	3.366 (out of 5)	0 (out of 45)

Table 19: The masking and swamping values for Right AFT-Loglogistic model $(n = 50)$

r	10							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.1052 (out of 10)	0.2527 (out of 90)	0.1195 (out of 10)	0.2641 (out of 90)	0.1145 (out of 10)	0.2493 (out of 90)	0.0621 (out of 10)	0.2681 (out of 90)
DG rob	9.8156 (out of 10)	0.0005 (out of 90)	9.8066 (out of 10)	0.0003 (out of 90)	9.8121 (out of 10)	0.0006 (out of 90)	7.8148 (out of 10)	0.0006 (out of 90)
DG	9.9993 (out of 10)	0 (out of 90)	9.9993 (out of 10)	0 (out of 90)	9.9993 (out of 10)	0 (out of 90)	9.8076 (out of 10)	0 (out of 90)

Table 20: The masking and swamping values for Right AFT-Loglogistic model $(n = 100)$

Masking decreases if the parameters characterizing remoteness of c-outliers increase. Let us discuss the results for small $(n = 20)$, medium $(n = 100)$, and large $(n = 1000)$ samples

r	100							
Method/ θ	0.001		0.005		0.02		0.05	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.0313 (out of 100)	0.18 (out of 900)	0.0351 (out of 100)	0.1898 (out of 900)	0.0267 (out of 100)	0.1935 (out of 900)	0.032 (out of 100)	0.1857 (out of 900)
DG rob	99.995 (out of 100)	0 (out of 900)	99.9958 (out of 100)	0 (out of 900)	99.995 (out of 100)	0 (out of 900)	99.9949 (out of 100)	0 (out of 900)
DG	99.9996 (out of 100)	0 (out of 900)	99.9999 (out of 100)	0 (out of 900)	99.9997 (out of 100)	0 (out of 900)	99.9996 (out of 100)	0 (out of 900)

Table 21: The masking and swamping values for Right AFT-Loglogistic model $(n = 1000)$

separately.

1) $n = 20$. Table 18 shows that when $\theta < 1$ the masking values (in terms of the numbers of nonidentified c-outliers) of the BP-method identifies about 0.75 outliers out of 2 and DG robust method identifies about 0.25 outliers out of 2. It is important to notice that for all θ masking values of BP and DG robust methods are not satisfactory having in mind that there is 2 outliers in the sample. Masking values for both methods decrease when θ increase, however, BP method masking values are smaller than DG robust method values. DG method without robust estimation shows worst results then BP and DG robust method. In this sample BP method has a heavier swamping effect, but smaller masking effect. Increasing θ values we notice that BP values varies about 0.2 when at that time DG robust swamping values also increase but are much smaller, DG method swamping values are equal 0. Therefore we can state that with small sample size BP method works better than DG robust or DG methods because have smaller masking effect.

2) $n = 100$. Table 20 shows that the masking values of the BP-method are much smaller then DG robust method with different θ. In this case DG robust and DG methods has heavy masking effect and DG robust method with $\theta = 1$ identifies only about 22% of outliers out of 10, DG without robust estimation identifies 2%, when BP method identifies more than 94% of outliers. In other hand DG robust and DG methods have less swamping effect impact than BP method with all θ, but all three methods swamping values are small enough. Considering all data it can be stated that BP method works better than DG robust and DG method with medium sample size.

3) $n = 1000$. Table 21 shows that DG robust and DG method have an enormous masking effect for all θ and less than 1% of outliers would be identified out of 100. This means DG robust and DG methods can not "catch" true outliers and can not be used for outlier identification for large samples. BP method for large sample size works correctly and have small masking effect on values. Swamping values in this event are not important, because masking value for DG robust and DG methods are near outlier value. In other hand DG robust and DG methods do not have swamping effect for all θ. For BP method swamping effect is minimum. Therefore DG robust and DG methods for large sample size are wrong, BP method works correctly.

Conclusion: BP method identifies outliers in AFT Loglogistic model more successfully than DG robust and DG methods.

4.3.2. Comparison between BP and Davies Gather for two-sided outlier search method

In this section was investigated medium size value ($n = 100$) outlier search method. This outlier search method was created by Section 3.6.3.

Let's compare BP, DG robust and DG methods with 10 outliers ($r_{left} = 5, r_{right} = 5$) and check how this three methods works for two-sided method with different truncated exponential distribution scale parameter θ. (see Table: 22).

r	10							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	0.7997 (out of 10)	0.1086 (out of 90)	0.8135 (out of 10)	0.1113 (out of 90)	0.7886 (out of 10)	0.1049 (out of 90)	0.4361 (out of 10)	0.112 (out of 90)
DG rob	9.8266 (out of 10)	0.0001 (out of 90)	9.8122 (out of 10)	0.0005 (out of 90)	9.8093 (out of 10)	0.0003 (out of 90)	7.823 (out of 10)	0.0011 (out of 90)
DG	9.9994 (out of 10)	0 (out of 90)	9.9988 (out of 10)	0 (out of 90)	9.999 (out of 10)	0 (out of 90)	9.8126 (out of 10)	0 (out of 90)

Table 22: The masking and swamping values for Two-Sided outlier search in AFT-Loglogistic model ($n = 100$)

Table 22 shows very similar results as we see in Section 4.3.1. Table 22 shows that the masking values of the BP-method are much smaller then DG robust method with different θ.

In this case DG robust and DG methods has heavy masking effect and BP robust method with $\theta = 1$ identifies only about 22% of outliers out of 10, DG without robust estimation identifies 2%, when BP method identifies more than 95% of outliers. In other hand DG robust and DG methods have less swamping effect impact than BP method with all θ, but all three methods swamping values are small enough. Considering all data it can be stated that BP method works better than DG robust and DG methods with medium sample size for two-sided outlier search case.

4.4. AFT-Lognormal regression model

4.4.1. Comparison between BP and Davies Gather for right outlier search method

In this section performance of BP, DG robust and DG methods for AFT Lognormal regression model was analyzed. Table 23 shows the values of the proportionsp_O of samples declared as having outliers. All three methods α values are near 0.05, which means that swamping effect should be minimal. Analyzing BP methods value, when $n = 1000$, we notice that shown value $\alpha = 0.0879$, this means that p_O is bigger than theoretical α.

Method	n=20	n=50	n=100	n=1000
	α	α	α	α
BP	0.0248	0.0287	0.0418	0.0879
DG rob	0.0424	0.0465	0.0477	0.0533
DG	0.0456	0.0525	0.0547	0.0551

Table 23: The proportions $\alpha = p_O$ values given that c-outliers are absent for right AFT-Lognormal regression model

Let's check how BP, DG robust and DG methods works with some outliers r and investigate swamping and masking effect. For $n = 20, 50, 100$ and 1000 the masking values are presented in Tables: 24, 25, 26, 27.

Masking decreases if the parameters characterizing remoteness of c-outliers increase. Let us discuss the results for small $(n = 20)$, medium $(n = 100)$, and large $(n = 1000)$ samples separately.

r	2							
Method/ θ	0.05		0.1		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	1.6493 (out of 2)	0.1091 (out of 18)	1.6403 (out of 2)	0.1092 (out of 18)	1.1486 (out of 2)	0.1223 (out of 18)	0.3526 (out of 2)	0.143 (out of 18)
DG rob	1.8432 (out of 2)	0.0077 (out of 18)	1.8301 (out of 2)	0.0087 (out of 18)	1.2754 (out of 2)	0.0107 (out of 18)	0.4413 (out of 2)	0.0161 (out of 18)
DG	1.9368 (out of 2)	0 (out of 18)	1.9366 (out of 2)	0 (out of 18)	1.5126 (out of 2)	0 (out of 18)	0.9094 (out of 2)	0 (out of 18)

Table 24: The masking and swamping values for Right AFT-Lognormal model ($n = 20$)

r	5							
Method/ θ	0.05		0.1		1		5	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	1.4628 (out of 5)	0.1747 (out of 45)	1.5351 (out of 5)	0.1766 (out of 45)	0.461 (out of 5)	0.1962 (out of 45)	0.0737 (out of 5)	0.198 (out of 45)
DG rob	4.8079 (out of 5)	0.0013 (out of 45)	4.8109 (out of 5)	0.001 (out of 45)	2.8318 (out of 5)	0.0027 (out of 45)	0.8491 (out of 5)	0.005 (out of 45)
DG	4.9747 (out of 5)	0 (out of 45)	4.9762 (out of 5)	0 (out of 45)	4.0189 (out of 5)	0 (out of 45)	2.4048 (out of 5)	0 (out of 45)

Table 25: The masking and swamping values for Right AFT-Lognormal model ($n = 50$)

r	10							
Method/ θ	0.001		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	1.3387 (out of 10)	0.1577 (out of 90)	1.3064 (out of 10)	0.1608 (out of 90)	1.3065 (out of 10)	0.1549 (out of 90)	0.3916 (out of 10)	0.1788 (out of 90)
DG rob	9.9087 (out of 10)	0.0002 (out of 90)	9.8977 (out of 10)	0.0001 (out of 90)	9.9056 (out of 10)	0 (out of 90)	5.606 (out of 10)	0.0003 (out of 90)
DG	9.9956 (out of 10)	0 (out of 90)	9.997 (out of 10)	0 (out of 90)	9.9972 (out of 10)	0 (out of 90)	8.3419 (out of 10)	0 (out of 90)

Table 26: The masking and swamping values for Right AFT-Lognormal model ($n = 100$)

1) $n = 20$. Table 24 shows that when $\theta < 1$ the masking values (in terms of the numbers of nonidentified c-outliers) of the BP-method identifies about 0.35 outliers out of 2 and DG

r	100							
Method/ θ	0.001		0.005		0.02		0.05	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	1.723 (out of 100)	0.0548 (out of 900)	1.7791 (out of 100)	0.0607 (out of 900)	1.8221 (out of 100)	0.0568 (out of 900)	1.6423 (out of 100)	0.0572 (out of 900)
DG rob	99.9992 (out of 100)	0 (out of 900)	99.9993 (out of 100)	0 (out of 900)	99.9988 (out of 100)	0 (out of 900)	99.9995 (out of 100)	0 (out of 900)
DG	100 (out of 100)	0 (out of 900)	100 (out of 100)	0 (out of 900)	99.9996 (out of 100)	0 (out of 900)	100 (out of 100)	0 (out of 900)

Table 27: The masking and swamping values for Right AFT-Lognormal model ($n = 1000$)

robust method identifies about 0.15 outliers out of 2 and DG method identifies about 0.06 outliers out of 2. It is important to notice that for all θ masking values of three methods are not satisfactory having in mind that there is 2 outliers in the sample. Masking values for three methods decrease when θ increase, however, BP method masking values are smaller than DG robust and DG methods values. In this sample BP method has a heavier swamping effect, but smaller masking effect. Increasing θ values we notice that BP swamping values varies about 0.14 when at that time DG robust swamping values also increase but are much smaller, DG method swamping values are 0. Therefore we can state that with small sample size BP methods works better than DG robust or DG method because have smaller masking effect. DG method without robust estimation with small sample sizes do not work correctly.

2) $n = 100$. Table 26 shows that the masking values of the BP-method are much smaller then DG robust method with different θ. In this case DG robust and DG methods has heavy masking effect, but DG robust method is better then DG method without robust estimation, because DG robust method has smaller masking effect. DG robust method with $\theta = 1$ identifies only about 45% of outliers out of 10, when BP method identifies about 96% of outliers. In other hand DG robust method has less swamping effect impact than BP method with all θ, but both of the methods swamping values are small enough. Considering all data it can be stated that BP method works better than DG robust and DG methods with medium sample size.

3) $n = 1000$. Table 27 shows that DG robust and DG method have an enormous masking effect for all θ and for DG robust method less than 1% of outliers would be identified out of

100. DG method without robust estimation can not define any outlier. This means DG robust and DG methods can not "catch" true outliers and can not be used for outlier identification for large samples. BP method for large sample size works correctly and have small masking effect on values. Swamping values in this event are not important, because masking values for DG robust and DG methods are near outlier value. In other hand DG robust and DG methods do not have swamping effect for all θ. For BP method swamping effect is minimum. Therefore DG robust and DG methods for large sample size are wrong, BP method works correctly.

Conclusion: BP method identifies outliers in AFT Lognormal model more successfully than DG robust and DG methods.

4.4.2. Comparison between BP and Davies Gather for two-sided outlier search method

In this section was investigated medium size value ($n = 100$) outlier search method. This outlier search method was created by Section 3.6.3.

Let's compare BP, DG robust and DG methods with 10 outliers ($r_{left} = 5, r_{right} = 5$) and check how this two methods works for two-sided method with different truncated exponential distribution scale parameter θ. (see Table: 28).

r	10							
Method/ θ	0.01		0.05		0.1		1	
	Masking	Swamping	Masking	Swamping	Masking	Swamping	Masking	Swamping
BP	3.6216 (out of 10)	0.0561 (out of 90)	3.5855 (out of 10)	0.0534 (out of 90)	3.6331 (out of 10)	0.0521 (out of 90)	1.3875 (out of 10)	0.0704 (out of 90)
DG rob	9.9068 (out of 10)	0.0001 (out of 90)	9.9094 (out of 10)	0 (out of 90)	9.9029 (out of 10)	0 (out of 90)	5.5856 (out of 10)	0.0005 (out of 90)
DG	9.9963 (out of 10)	0 (out of 90)	9.9968 (out of 10)	0 (out of 90)	9.9961 (out of 10)	0 (out of 90)	8.3298 (out of 10)	0 (out of 90)

Table 28: The masking and swamping values for two sided outlier search in AFT-Lognormal model ($n = 100$)

Masking decreases if the parameters characterizing remoteness of c-outliers increase. Table 28 shows that the masking values of the BP-method are much smaller then DG robust method with different θ. In this case, DG robust and DG methods has heavy masking effect, but DG robust method is better then DG method without robust estimation, because DG

robust method has smaller masking effect. DG robust method with $\theta = 1$ identifies only about 45% of outliers out of 10, when BP method identifies about 86% of outliers. In other hand, DG robust method has less swamping effect impact than BP method with all θ, but both of the methods swamping values are small enough. Considering all data it can be stated that BP method works better than DG robust and DG methods with medium sample size for two-sided outlier search. In this section we got very similar results as 4.4 Section.

5. Real data example

In this section real data set example (from [28] and [35]) using BP and DG robust method was investigated.

5.1. Wayne Nelson data set

Wayne Nelson [28] studied failure times of 76 units of insulating fluids. Testing was performed at various constant elevated voltages ranging from 26 to 38 kilovolts (kV). The number of batches assigned to the different voltage levels were $3, 5, 11, 15, 19, 15$, and 8, respectively. This experiment was run long enough to observe the failures of all items. The voltage levels v_i, the numbers of items tested under the same voltage level, and the failure times T_i were shown in scatter plot (see Figure 1). To check how correctly BP and DG robust method works, was created three outliers (1 left and 2 right)

Figure 1 suggests transformation of variables. The voltage levels v_i and the failure times T_i were transformed by natural logarithm (see Figure 2)

The AFT-Weibull regression model is:

$$log(T_i) = 65.722 - 18.112 log(v_i)$$

Using BP method all three outliers were found, using DG robust method one left and one right (from two) were found (see Figure 3).

After removal of outliers AIC values for three different AFT models (AFT-Weibull, AFT-lognormal and AFT-lognormal) were compared. Table 29) shows that the AFT-Weibull model is the most appropriate (AFT-Weibull has smallest AIC value). The chi-square good-

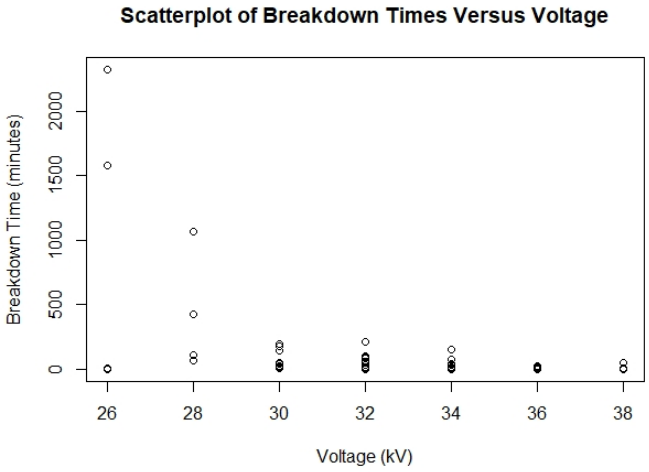

Figure 1: Scatterplot of Wayne Nelson data

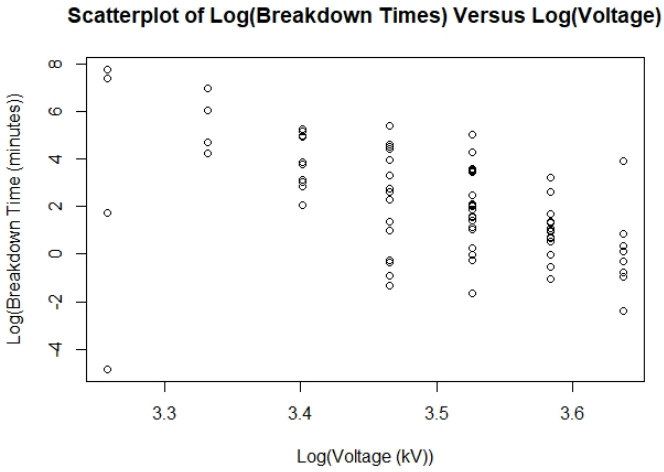

Figure 2: Scatterplot of Wayne Nelson logarithmic data

ness of fit test for AFT-Weibull model given by Bagdonavičius-Levulienė-Nikulin [2] accepts the AFT-Weibull model ($p-value = 0.4604$).

Distribution	AIC
AFT-Weibull	607.0694
AFT-Loglogistic	612.8227
AFT-Lognormal	613.3748

Table 29: AFT regressions AIC values for Wayne Nelson data

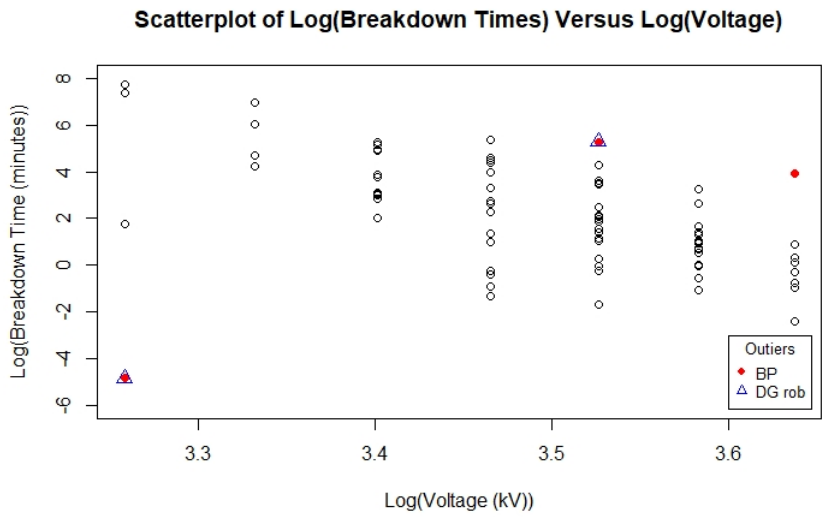

Figure 3: Scatterplot of Wayne Nelson data with outliers (found by BP and DG robust methods)

5.2. Small-leaved lime trees grown in Russia

In this section data from small-leaved lime trees grown in Russia [35] were investigated. This dataset was taken from `R` package `GLMsData`. A data frame consists of 385 observations with 4 different attributes: foliage, DBH, age and origin. To the outliers search we used data from trees of a natural origin (origin = Natural). The foliage (the foliage biomass, in kg (oven dried matter)) and DBH (the tree diameter, at breast height, in cm) are shown in scatter plot (see: Figure 4). In this case age was not investigated.

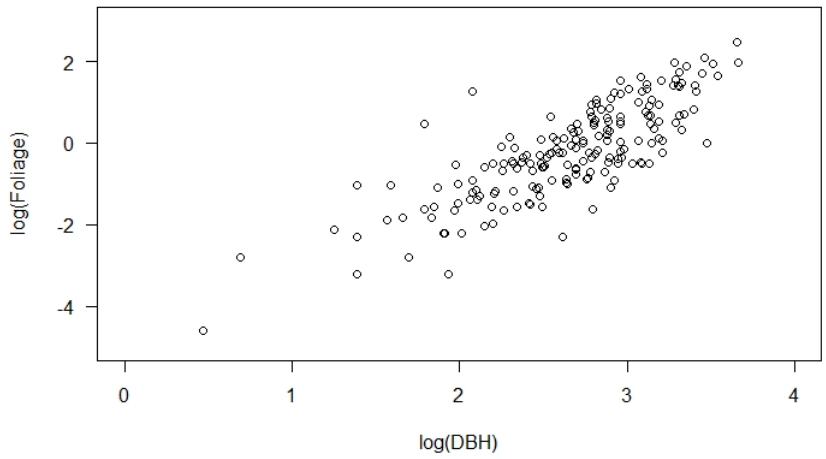

Figure 4: Scatterplot of log(Foliage) and log(DBH)

Using AFT-Lognormal model BP method two outliers (red points) were found, using DG robust method one outlier (blue triangle) was found (see Figure 5).

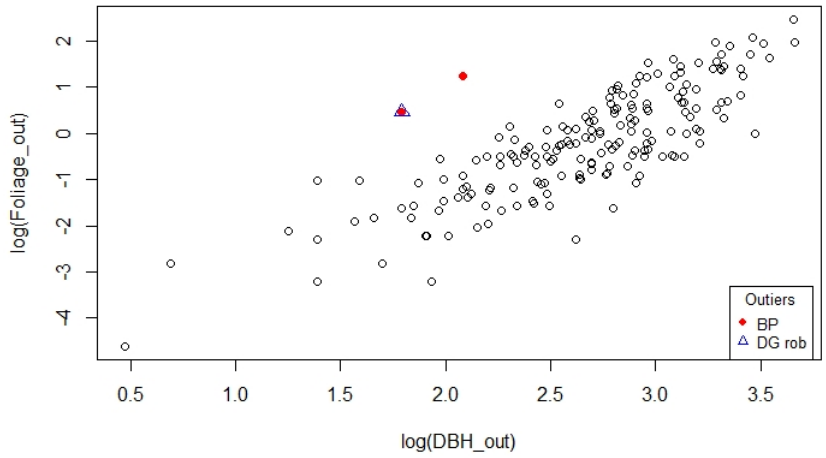

Figure 5: Scatterplot of log(Foliage) and log(DBH) with outliers (found by BP and DG robust methods)

After removal of outliers AIC values for three different AFT models (AFT-Weibull, AFT-lognormal and AFT-lognormal) were compared. Table 30 shows that the AFT-Lognormal model is the most appropriate (AFT-Lognormal has smallest AIC value).

Distribution	AIC
AFT-Weibull	286.5589
AFT-Loglogistic	291.9761
AFT-Lognormal	284.5503

Table 30: AFT regressions AIC values for Small-leaved lime trees data

6. Conclusions

In many situations, the BP outlier identification method has superior performance as compared to existing methods for accelerated failure time regression models. BP and DG robust methods in all situations show better results then DG method without robust estimation. The BP method is based on an asymptotic result, so it should not be applied for samples of very small size $n \leq 15$. Two real life examples confirm suggestion that BP method identifies outliers better than DG robust method.

References

[1] Vilijandas Bagdonavičius and Linas Petkevičius. “A new multiple outliers identification method in linear regression”. In: *Metrika* 83.3 (2020), pp. 275–296.

[2] Vilijandas B Bagdonavičius, Rūta J Levuliene, and Mikhail S Nikulin. “Chi-squared goodness-of-fit tests for parametric accelerated failure time models”. In: *Communications in Statistics-Theory and Methods* 42.15 (2013), pp. 2768–2785.

[3] Vilijandas Bagdonavičius and Linas Petkevičius. “Multiple outlier detection tests for parametric models”. In: *arXiv preprint arXiv:1910.10426* (2019).

[4] Vic Barnett and Toby Lewis. “Outliers in statistical data”. In: *Wiley* (1974).

[5] David A Belsley, Edwin Kuh, and Roy E Welsch. *Regression diagnostics: Identifying influential data and sources of collinearity*. Vol. 571. John Wiley, 1980.

[6] Nedret Billor, Ali S Hadi, and Paul F Velleman. “BACON: blocked adaptive computationally efficient outlier nominators”. In: *Computational statistics & data analysis* 34.3 (2000), pp. 279–298.

[7] LN Bol’shev and M Ubaidullaeva. “Chauvenet’s test in the classical theory of errors”. In: *Theory of Probability & Its Applications* 19.4 (1975), pp. 683–692.

[8] Samprit Chatterjee and Ali S Hadi. *Regression analysis by example*. John Wiley & Sons, 2015.

[9] MS Chikkagoudar and SH Kunchur. “Distributions of test statistics for multiple outliers in exponential samples”. In: *Communications in Statistics-theory and Methods* 12.18 (1983), pp. 2127–2142.

[10] R Dennis Cook. “Detection of influential observation in linear regression”. In: *Technometrics* 19.1 (1977), pp. 15–18.

[11] R Dennis Cook. “Influential observations in linear regression”. In: *Journal of the American Statistical Association* 74.365 (1979), pp. 169–174.

[12] Laurie Davies and Ursula Gather. “The identification of multiple outliers”. In: *Journal of the American Statistical Association* 88.423 (1993), pp. 782–792.

[13] Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*. Springer Science & Business Media, 2007.

[14] John Fox. *Regression diagnostics: An introduction*. Vol. 79. Sage Publications, 1991.

[15] Frank E Grubbs et al. "Sample criteria for testing outlying observations". In: *The Annals of Mathematical Statistics* 21.1 (1950), pp. 27–58.

[16] Ali S Hadi. "A new measure of overall potential influence in linear regression". In: *Computational Statistics & Data Analysis* 14.1 (1992), pp. 1–27.

[17] Douglas M Hawkins. *Identification of outliers*. Vol. 11. Springer, 1980.

[18] DG Kabe. "Testing outliers from an exponential population". In: *Metrika* 15.1 (1970), pp. 15–18.

[19] AC Kimber. "Testing upper and lower outlier paris in gamma samples". In: *Communications in Statistics-Simulation and Computation* 17.3 (1988), pp. 1055–1072.

[20] AC Kimber. "Tests for many outliers in an exponential sample". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 31.3 (1982), pp. 263–271.

[21] S Lalitha and Nirpeksh Kumar. "Multiple outlier test for upper outliers in an exponential sample". In: *Journal of applied statistics* 39.6 (2012), pp. 1323–1330.

[22] Annick M Leroy and Peter J Rousseeuw. "Robust regression and outlier detection". In: *rrod* (1987).

[23] T Lewis and NRJ Fieller. "A recursive algorithm for null distributions for outliers: I. Gamma samples". In: *Technometrics* 21.3 (1979), pp. 371–376.

[24] J Likeš. "Distribution of Dixon's statistics in the case of an exponential population". In: *Metrika* 11.1 (1967), pp. 46–54.

[25] Chien-Tai Lin and N Balakrishnan. "Exact computation of the null distribution of a test for multiple outliers in an exponential sample". In: *Computational statistics & data analysis* 53.9 (2009), pp. 3281–3290.

[26] Chien-tai Lin and N Balakrishnan. "Tests for multiple outliers in an exponential sample". In: *Communications in Statistics-Simulation and Computation* 43.4 (2014), pp. 706–722.

[27] Martin Maechler et al. “robustbase: Basic Robust Statistics; 2018”. In: *URL http://robustbase. r-forge. r-project. org/. R package version 0.92-7.[p310]* ().

[28] Wayne Nelson. “Graphical analysis of accelerated life test data with the inverse power law model”. In: *IEEE Transactions on Reliability* 21.1 (1972), pp. 2–11.

[29] Abdul Awal Md Nurunnabi and Honghua Dai. “Robust-diagnostic regression: a prelude for inducing reliable knowledge from regression”. In: *Reliable knowledge discovery*. Springer, 2012, pp. 69–92.

[30] Daniel Peña. “A new statistic for influence in linear regression”. In: *Technometrics* 47.1 (2005), pp. 1–12.

[31] AHM Rahmatullah Imon. “Identifying multiple influential observations in linear regression”. In: *Journal of Applied statistics* 32.9 (2005), pp. 929–946.

[32] Marco Riani and Anthony C Atkinson. “Robust diagnostic data analysis: transformations in regression”. In: *Technometrics* 42.4 (2000), pp. 384–394.

[33] Bernard Rosner. “On the detection of many outliers”. In: *Technometrics* 17.2 (1975), pp. 221–227.

[34] Peter J Rousseeuw and Bert C Van Zomeren. “Unmasking multivariate outliers and leverage points”. In: *Journal of the American Statistical association* 85.411 (1990), pp. 633–639.

[35] D Schepaschenko et al. “A database of forest biomass structure for Eurasia”. In: (2017).

[36] Gary L Tietjen and Roger H Moore. “Some Grubbs-type statistics for the detection of several outliers”. In: *Technometrics* 14.3 (1972), pp. 583–597.

[37] Valentin Todorov, Peter Filzmoser, et al. “An object-oriented framework for robust multivariate analysis”. In: (2009).

[38] Roy E Welsch and Edwin Kuh. *Linear regression diagnostics*. Tech. rep. National Bureau of Economic Research, 1977.

[39] Dixon WJ. “Analysis of Extreme Values”. In: *The Annals of Mathematical Statistics* 21.4 (1950), pp. 488–506.

[40] Aïcha Zerbet. “Statistical Tests for Normal Family in Presence of Outlying Observations”. In: *Goodness-of-Fit Tests and Model Validity*. Springer, 2002, pp. 57–64.

[41] Aicha Zerbet and Mikhail Nikulin. “A new statistic for detecting outliers in exponential case”. In: *Communications in Statistics-theory and Methods* 32.3 (2003), pp. 573–583.