

**Robust Testing  
Between Two Asymmetric  
Distributions:  
A Simulation Study**

by

Diego Kuonen

`kuonen@dma.epfl.ch`

Department of Mathematics  
Swiss Federal Institute of Technology  
Lausanne, CH

Supervised by Prof. A. Marazzi, IUMSP, Lausanne

June 1997

## Abstract

In this paper a robust test statistic for testing between two income distributions is analysed. Victoria-Feser (1996) proposed a robust version of the classical test based on the comparison of the maximised likelihood functions for the two models (Cox 1961, 1962). She showed its robustness properties as well as other properties. She also proposed an algorithm for the computation of the robust Cox-type test statistic. This algorithm is initialised by taking the classical maximum likelihood estimates of the parameters. But would it not be more appropriate to take a robust estimator to estimate the parameters of the models? The aim of this paper consists in answering this question in use of simulated examples. S-PLUS functions for computing the described procedures are made available.

**Key words:** M-estimators, model choice, robust tests, Exponential distribution, Pareto distribution, Lognormal distribution, Weibull distribution, Gamma distribution

# Table of Contents

<b>1</b>	<b>Robust Cox-Type Statistics</b>	<b>1</b>
1.1	The Victoria-Feser Statistic . . . . .	1
1.2	The Marazzi Statistic . . . . .	3
1.3	An Empirical Comparison Between the Approaches . . . . .	5
1.4	The Tuning Parameters . . . . .	7
<b>2</b>	<b>Simulation Study</b>	<b>8</b>
2.1	Preliminary . . . . .	8
2.2	Simulations . . . . .	11
2.2.1	Pareto against Exponential . . . . .	14
2.2.2	Exponential against Pareto . . . . .	17
2.2.3	Gamma against Weibull . . . . .	19
2.2.4	Gamma against Lognormal . . . . .	21
<b>3</b>	<b>Outstanding Questions</b>	<b>22</b>
<b>4</b>	<b>Summary and Conclusions</b>	<b>24</b>
<b>A</b>	<b>Appendix</b>	<b>26</b>
A.1	S-PLUS Functions for the Pareto Distribution . . . . .	26
A.2	S-PLUS Functions for the Computation of the Statistics . . . . .	28
	<b>References</b>	<b>33</b>

## List of Figures

1	Histogram of lengths of stay of patients hospitalised in Switzerland . . . . .	5
2	Histogram of lengths of stay with maximum likelihood estimates . . . . .	6
3	Histogram of lengths of stay with $B_s^p$ - estimates . . . . .	7
4	Histogram of 5000 test statistics . . . . .	8
5	Asymptotic and simulated critical value - Gamma . . . . .	9
6	Contaminated Exponential and Pareto distributions . . . . .	12
7	Contaminated Gamma, Weibull and Lognormal distributions . . . . .	12
8	Gamma and Weibull distribution . . . . .	20
9	Independence of the Cox statistic - Exponential against Pareto . . . . .	23
10	Independence of the Cox statistic - Gamma against Lognormal . . . . .	23
11	Pareto samples generated by means of the S-PLUS function <code>rpareto</code> . . . . .	27

## List of Tables

1	Lengths of stay of patients hospitalised in Switzerland . . . . .	5
2	Asymptotic and simulated critical values - Gamma . . . . .	9
3	Asymptotic and simulated critical values - Weibull . . . . .	10
4	Probability of rejection - Gamma . . . . .	10
5	Probability of rejection - Weibull . . . . .	10
6	Parameter choice for the contamination. . . . .	11
7	Comparison of the contaminations . . . . .	13
8	Attained significance level - Pareto against Exponential . . . . .	15
9	Attained power - Pareto against Exponential . . . . .	15
10	Simulated Pareto percentiles - Pareto against Exponential . . . . .	16
11	Simulated Exponential percentiles - Pareto against Exponential . . . . .	16
12	Attained significance level - Exponential against Pareto . . . . .	17
13	Attained power - Exponential against Pareto . . . . .	18
14	Simulated Pareto percentiles - Exponential against Pareto . . . . .	18
15	Attained significance level - Gamma against Weibull . . . . .	19
16	Attained power - Gamma against Weibull . . . . .	20
17	Attained significance level - Gamma against Lognormal . . . . .	21

# 1 Robust Cox-Type Statistics

Cox (1961, 1962) proposed a test statistic to choose between two models. The aim of this test is to choose the model which best represents the data observed. Victoria-Feser (1996) showed that this test statistic is not robust and proposed a robust version of the Cox test statistic which is initialised by taking the classical maximum likelihood estimates of the parameters. This new test was obtained by considering the Cox test as a Lagrange multiplier or score test and in applying the results on robust parametric tests.

In this section we first recapitulate the robust test statistic proposed by Victoria-Feser, and in a second time we define a new test statistic which is initialised by taking the robust estimates of the parameters. It can be seen as a generalisation of the one proposed by Victoria-Feser (1996).

## 1.1 The Victoria-Feser Statistic

Denote the vector random variable to be observed by  $Y$ . Let  $y_1, \dots, y_n$  be a sample of  $Y$ , and let  $H_0$  be the hypothesis that their distribution function is  $F_\alpha^0$  with corresponding density  $f^0(y; \alpha)$ , and let  $H_1$  be the hypothesis that their distribution function is  $F_\beta^1$  with density  $f^1(y; \beta)$ . We assume that the hypotheses  $H_0$  and  $H_1$  are separate in the sense that an arbitrary simple hypothesis in  $H_0$  cannot be obtained as a limit of simple hypotheses in  $H_1$  (Cox 1961). This contrasts with the usual situation in hypothesis testing. We consider the special case where  $\alpha$  and  $\beta$  are parameter vectors. So let be  $\alpha = (\theta_1, \theta_2)$  and  $\beta = (\nu_1, \nu_2)$ . The log likelihood functions are  $L_0(y; (\theta_1, \theta_2)) = \log f^0(y; (\theta_1, \theta_2))$  and  $L_1(y; (\nu_1, \nu_2)) = \log f^1(y; (\nu_1, \nu_2))$ , and their difference is denoted by  $L(y; \theta_1, \theta_2, \nu_1, \nu_2) = L_0(y; (\theta_1, \theta_2)) - L_1(y; (\nu_1, \nu_2))$ . Cox (1961) based his test on the following statistic

$$\mathbf{U}_{\text{cox}} = \frac{1}{n} \sum_{i=1}^n L(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) - \int L(y; \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) f^0(y; \hat{\nu}_1, \hat{\nu}_2) dy \quad (1)$$

where  $\hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1$  and  $\hat{\nu}_2$  are the maximum likelihood estimates of  $\theta_1, \theta_2, \nu_1$  and  $\nu_2$ , under  $H_0$ . Cox (1962) showed that, under  $H_0$ , the statistic  $\mathbf{U}_{\text{cox}}$  is asymptotically normally distributed with mean 0 and with variance  $\text{Var}(F_\alpha^0)$  (for details see Cox, 1962, or Victoria-Feser, 1996). This test rejects  $H_0$  if  $|n^{1/2} \mathbf{U}_{\text{cox}}| > \kappa_\omega^*$  where  $\kappa_\omega^* = \phi^{-1}(1 - \omega/2)$ . This asymptotic distribution theory will still hold if instead of a maximum likelihood estimator any asymptotically equivalent method of estimation is used. Atkinson (1970) showed that this statistic can be interpreted as a Lagrange multiplier or score test. In fact in constructing a comprehensive model (see Victoria-Feser, 1996) the Lagrange multiplier or score test corresponds in testing between two simple hypotheses. Indeed, the density of this compound model is denoted by  $f^*(y; (\theta_1, \theta_2), (\nu_1, \nu_2), \lambda)$  which, under  $H_0 : \lambda = 1$ , equals to  $f^0(y; (\theta_1, \theta_2))$ , and under  $H_1 : \lambda = 0$ , equals to  $f^1(y; (\nu_1, \nu_2))$ . In considering this compound model Victoria-Feser (1996) based her robust version of the Cox statistic on the results of Heritier and Ronchetti (1994) who proposed a robust score test statistic

based on M-estimators. The score function, under  $H_0 : \lambda = 1$ , is given by

$$s^*(y; \theta_1, \theta_2, \nu_1, \nu_2) = \begin{pmatrix} s^0(y; \theta_1, \theta_2) \\ L(y; \theta_1, \theta_2, \nu_1, \nu_2) - \int L(y; \theta_1, \theta_2, \nu_1, \nu_2) f^0(y; \theta_1, \theta_2) dy \end{pmatrix}$$

where  $s^0(y; \theta_1, \theta_2)$  is the vector of the score functions under  $F_{(\theta_1, \theta_2)}^0$ , *i.e.*

$$s^0(y; \theta_1, \theta_2) = \begin{pmatrix} s_1(y; \theta_1, \theta_2) \\ s_2(y; \theta_1, \theta_2) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \theta_1} L_0(y; (\theta_1, \theta_2)) \\ \frac{\partial}{\partial \theta_2} L_0(y; (\theta_1, \theta_2)) \end{pmatrix}$$

Let  $s_3(y; \theta_1, \theta_2, \nu_1, \nu_2)$  be  $L(y; \theta_1, \theta_2, \nu_1, \nu_2) - \int L(y; \theta_1, \theta_2, \nu_1, \nu_2) f^0(y; \theta_1, \theta_2) dy$ . As the  $\psi$ -function of a M-estimator is defined by its influence function, and as the influence function of a maximum likelihood estimator is of the form  $J(\hat{\theta}_1, \hat{\theta}_2)^{-1} s^0(y; \theta_1, \theta_2)$ , where  $J(\hat{\theta}_1, \hat{\theta}_2)^{-1}$  is the inverse of the Fisher information matrix, the first two components of the  $\psi$ -function are determined easily. Hence for a given bound  $b$ , the robust test statistic is based on the following  $\psi$ -function

$$\psi_b^*(y; \theta_1, \theta_2, \nu_1, \nu_2) = \begin{pmatrix} j_{11} s_1(y; \theta_1, \theta_2) + j_{12} s_2(y; \theta_1, \theta_2) \\ j_{12} s_1(y; \theta_1, \theta_2) + j_{22} s_2(y; \theta_1, \theta_2) \\ h_b \{ a_{11} s_1(y; \theta_1, \theta_2) + a_{21} s_2(y; \theta_1, \theta_2) + a_{22} (s_3(y; \theta_1, \theta_2, \nu_1, \nu_2) - c) \} \end{pmatrix}$$

where  $h_b\{x\} = x \min(1; b/|x|)$  is Huber's function with tuning constant  $b$  and  $j_{rs}$  ( $r = 1, 2$  and  $s = 1, 2$ ) are the elements of the symmetric  $2 \times 2$  matrix  $J(\theta_1, \theta_2)^{-1}$ . The tuning constant is used to regulate the degree of robustness, and the centring parameter  $c$  is used to achieve Fisher consistency. Moreover  $c$  and the lower triangular  $2 \times 2$  matrix  $A$  are determined as solutions of

$$\begin{aligned} \int \psi_b^*(y; \theta_1, \theta_2, \nu_1, \nu_2) f^0(y; \theta_1, \theta_2) dy &= 0 \\ \int \psi_b^*(y; \theta_1, \theta_2, \nu_1, \nu_2) \psi_b^*(y; \theta_1, \theta_2, \nu_1, \nu_2)^T f^0(y; \theta_1, \theta_2) dy &= I \end{aligned}$$

As this test relies on M-estimators defined by

$$\sum_{i=1}^n \psi_b^*(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) = 0$$

one obtains from the first two components of  $\psi_b^*(y; \theta_1, \theta_2, \nu_1, \nu_2)$  the following equations

$$\sum_{i=1}^n s_1(y_i; \hat{\theta}_1, \hat{\theta}_2) = 0 \quad \text{and} \quad \sum_{i=1}^n s_2(y_i; \hat{\theta}_1, \hat{\theta}_2) = 0$$

One remarks that the solutions of this equations are the classical maximum likelihood estimates of the parameters. The Victoria-Feser test statistic (1996) is finally given by

$$\mathbf{U}_{\mathbf{VF}} = \frac{1}{n} \sum_{i=1}^n h_b \{ a_{11} s_1(y_i; \hat{\theta}_1, \hat{\theta}_2) + a_{21} s_2(y_i; \hat{\theta}_1, \hat{\theta}_2) + a_{22} (s_3(y_i; \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) - c) \} \quad (2)$$

where  $\hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1$  and  $\hat{\nu}_2$  are the maximum likelihood estimates of  $\theta_1, \theta_2, \nu_1$  and  $\nu_2$ . The asymptotic normality of  $n^{1/2} \mathbf{U}_{\mathbf{VF}}$  under  $H_0$  is proven in Heritier and Ronchetti (1994).

**Remark:** In taking the maximum likelihood estimates of the parameters this test statistic is an application of the optimal standardised  $B$ -robust estimator (called  $B_s$ -estimator) proposed by Hampel (1986). Among the proposals of Hampel (1986) is also the optimal standardised  $B$ -robust estimator for partitioned parameters (called  $B_s^p$ -estimator). Both estimators are studied and compared, in the case of the Gamma distribution, by Marazzi and Ruffieux (1996a), who conclude that the statistical performances of the  $B_s$ - and the  $B_s^p$ -estimator are similar. For practical use the  $B_s^p$ -estimator should be preferred because it is computationally simpler and more reliable.

## 1.2 The Marazzi Statistic

In many statistical models the parameter splits up in a natural way into a main part and a nuisance part. Moreover it is intuitively clear that the maximal bias of the main part in a contamination model (as it is usually the case with real data) depends on the maximal bias of the nuisance part. Therefore it would be interesting to treat both parameters in a symmetric manner. This is exactly what the optimal standardised  $B$ -robust estimator for partitioned parameters, *i.e.* the  $B_s^p$ -estimator, proposed by Hampel (1986) does. For this reason we define  $\psi_{\mathbf{b}}^*(y; \theta_1, \theta_2, \nu_1, \nu_2)$  by

$$\begin{pmatrix} h_{b_1} \{a_{11}(s_1(y; \theta_1, \theta_2) - c_1)\} \\ h_{b_2} \{a_{21}(s_1(y; \theta_1, \theta_2) - c_1) + a_{22}(s_2(y; \theta_1, \theta_2) - c_2)\} \\ h_{b_3} \{a_{31}(s_1(y; \theta_1, \theta_2) - c_1) + a_{32}(s_2(y; \theta_1, \theta_2) - c_2) + a_{33}(s_3(y; \theta_1, \theta_2, \nu_1, \nu_2) - c_3)\} \end{pmatrix}$$

where  $\mathbf{b} = (b_1, b_2, b_3)$  is the vector of the tuning parameters for Huber's functions,  $A$  is a  $3 \times 3$  nonsingular lower triangular matrix defined by

$$\int \psi_{\mathbf{b}}^*(y; \theta_1, \theta_2, \nu_1, \nu_2) \psi_{\mathbf{b}}^*(y; \theta_1, \theta_2, \nu_1, \nu_2)^T f^0(y; \theta_1, \theta_2) dy = I$$

and the vector  $\mathbf{c} = (c_1, c_2, c_3)$  is defined by

$$\int \psi_{\mathbf{b}}^*(y; \theta_1, \theta_2, \nu_1, \nu_2) f^0(y; \theta_1, \theta_2) dy = 0$$

The matrix  $A$  orthogonalizes the parameters and the vector  $\mathbf{c}$  is a correction of the bias which is used to achieve Fisher consistency. At  $b_1 = b_2 = b_3 = \infty$  we have the classical non-robust statistic proposed by Cox (1961, 1962). For simplification, let us write

$$\psi_{\mathbf{b}}^*(y; \theta_1, \theta_2, \nu_1, \nu_2) = \begin{pmatrix} \psi_1(y, \theta_1, \theta_2) \\ \psi_2(y, \theta_1, \theta_2) \\ \psi_3(y, \theta_1, \theta_2, \nu_1, \nu_2) \end{pmatrix}$$

The  $B_s^p$ -estimates of the parameters are, for given  $b_1$  and  $b_2$ , defined as solutions of

$$\sum_{i=1}^n \psi_1(y_i, \theta_1, \theta_2) = \sum_{i=1}^n h_{b_1} \{a_{11}(s_1(y_i; \theta_1, \theta_2) - c_1)\} = 0 \quad (3)$$

and

$$\sum_{i=1}^n \psi_2(y, \theta_1, \theta_2) = \sum_{i=1}^n h_{b_2} \{a_{21}(s_1(y_i; \theta_1, \theta_2) - c_1) + a_{22}(s_2(y_i; \theta_1, \theta_2) - c_2)\} = 0 \quad (4)$$

where  $a_{11}, a_{21}, a_{22}, c_1$  and  $c_2$  are solutions of

$$\int \psi_1(y, \theta_1, \theta_2) \psi_1(y, \theta_1, \theta_2) f^0(y; (\theta_1, \theta_2)) dy = 1 \quad (5)$$

$$\int \psi_1(y, \theta_1, \theta_2) \psi_2(y, \theta_1, \theta_2) f^0(y; (\theta_1, \theta_2)) dy = 0 \quad (6)$$

$$\int \psi_2(y, \theta_1, \theta_2) \psi_2(y, \theta_1, \theta_2) f^0(y; (\theta_1, \theta_2)) dy = 1 \quad (7)$$

$$\int \psi_1(y, \theta_1, \theta_2) f^0(y; (\theta_1, \theta_2)) dy = 0 \quad (8)$$

$$\int \psi_2(y, \theta_1, \theta_2) f^0(y; (\theta_1, \theta_2)) dy = 0 \quad (9)$$

One notices that the  $B_s^p$ -estimates are a modification of the maximum likelihood estimates, where the corrected likelihood scores are truncated at  $\pm b_1$ , and  $\pm b_2$ , respectively. We define, for a given vector of tuning parameters  $\mathbf{b}$ , the following algorithm for the computation of the robust Cox-Type test statistic (the Marazzi statistic).

### The Computational Algorithm

---

**Step 1:** Compute the  $B_s^p$ -estimators  $(\hat{\theta}_1, \hat{\theta}_2)$  and  $(\hat{\nu}_1, \hat{\nu}_2)$  in use of equations (3) and (4). Compute  $(a_{11}, a_{21}, a_{22}, c_1, c_2)$  in use of equations (5)-(9).

**Step 2:** Solve for  $c_3, a_{31}, a_{32}, a_{33}$ :

$$\int \psi_3(y, \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) f^0(y; (\hat{\theta}_1, \hat{\theta}_2)) dy = 0 \quad (10)$$

$$\begin{cases} \int \psi_1(y, \hat{\theta}_1, \hat{\theta}_2) \psi_3(y, \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) f^0(y; (\hat{\theta}_1, \hat{\theta}_2)) dy = 0 \\ \int \psi_2(y, \hat{\theta}_1, \hat{\theta}_2) \psi_3(y, \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) f^0(y; (\hat{\theta}_1, \hat{\theta}_2)) dy = 0 \\ \int \psi_3(y, \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) \psi_3(y, \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) f^0(y; (\hat{\theta}_1, \hat{\theta}_2)) dy = 1 \end{cases} \quad (11)$$

From (10) we get  $c_3$ , and equations (11) lead to  $a_{31}, a_{32}$  and  $a_{33}$ .

**Step 3:** Compute

$$\mathbf{U}_M = \frac{1}{n} \sum_{i=1}^n \psi_3(y_i, \hat{\theta}_1, \hat{\theta}_2, \hat{\nu}_1, \hat{\nu}_2) \quad (12)$$

**Step 4:** At the nominal level  $\omega$ , accept  $H_0$  if  $|n^{1/2} \mathbf{U}_M| < \phi^{-1}(1 - \omega/2)$ .

---

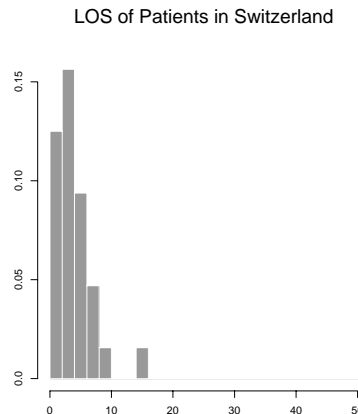
### 1.3 An Empirical Comparison Between the Approaches

In this section real data is analysed in order to know whether it is more appropriate to take the  $B_s^p$ -estimates of the parameters instead of the classical maximum likelihood estimates, as proposed by Victoria-Feser (1996). The data to be treated in this section is taken from Marazzi (1997a). The original data is summarised in table 1.

LOS	1	2	3	4	5	6	7	8	9	16	115	198	374
Freq.	2	6	5	5	4	2	2	1	1	1	1	1	1

**Table 1** Frequency distribution (Freq.) of lengths of stay (LOS) in days of patients hospitalised in Switzerland during 1988 for certain disorders of the nervous system.

Figure 1 shows the histogram of 32 lengths of stay (LOS) in days of patients hospitalised in Switzerland during 1988 for certain disorders of the nervous system. Note that the display is truncated at LOS=50.

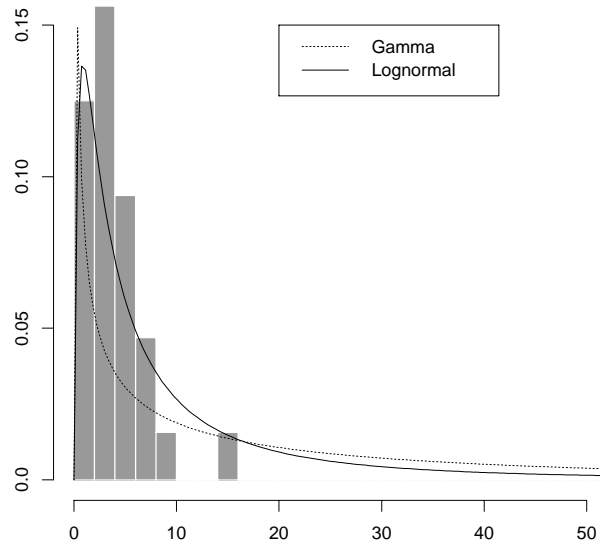


**Figure 1** Histogram of 32 lengths of stay (LOS) in days of patients hospitalised in Switzerland during 1988 for certain disorders of the nervous system (taken from Marazzi, 1997).

As the LOS represent a rounded continuous variable, we are not bothered by the use of continuous models. Suppose that we want to test the hypothesis  $H_0$ : ‘Lognormal’ against  $H_1$ : ‘Gamma’. In computing the maximum likelihood estimates of the parameters we are able to superpose the estimated density curve on the histogram. Figure 2 shows the two corresponding densities in taking the Gamma distribution (dashed line) and in taking the Lognormal distribution (solid line), respectively. In considering the situation of figure 2, one would intuitively prefer the Lognormal model. In computing the Victoria-Feser

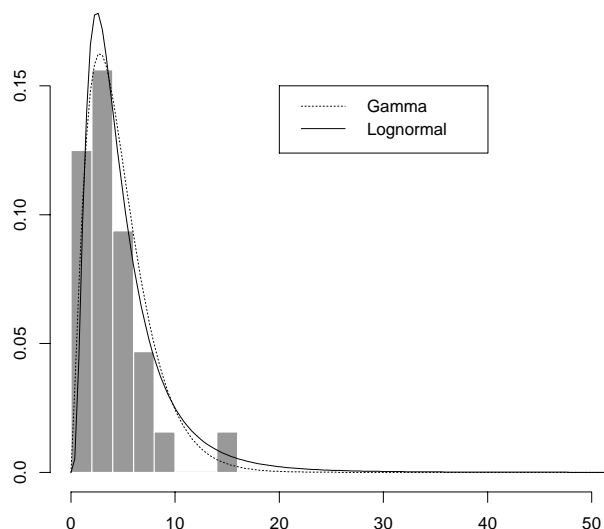
statistic defined by equation (2),  $|n^{1/2}\mathbf{U}_{VF}|$  equals 2.932 with corresponding attained significance level (ASL) of 0.5%. Hence we have to reject  $H_0$  in favour of  $H_1$ , *i.e.* we accept the Gamma distribution. This is clearly in opposition with figure 2, where the Lognormal distribution seems to fit the data better. Moreover, figure 3 has been obtained by robustly fitting (in taking the  $B_s^p$ -estimates) the Gamma respectively the Lognormal distribution. The Marazzi statistic defined by equation (12) provides 0.340 for  $|n^{1/2}\mathbf{U}_M|$  with ASL of 37,6%. Hence there is definitely no evidence against  $H_0$ . In conclusion we can say that in taking the robust estimates for the parameters we confirm the result from figures 2 and 3. Finally, we also calculated the original Cox statistic  $\mathbf{U}_{cox}$  defined by equation (1): we obtained 4.304 with ASL equal to 0,003%. Hence the choice of taking the  $B_s^p$ -estimates of the parameters instead of the classical maximum likelihood estimates seems to be justified.

With Maximum Likelihood Estimates



**Figure 2** Histogram of 32 lengths of stay (LOS) in days of patients hospitalised in Switzerland during 1988 and the adjusted densities in taking the maximum likelihood estimates of the parameters.

## With M-estimates



**Figure 3** Histogram of 32 lengths of stay (LOS) in days of patients hospitalised in Switzerland during 1988 and the adjusted densities in taking the  $B_s^p$ -estimates of the parameters.

## 1.4 The Tuning Parameters

The  $B_s^p$ -estimates of the parameters depend on the tuning constants  $b_1$  and  $b_2$  (see equations (3) and (4)) that must be chosen by the user. The most common rule for determining the tuning constants  $b_1$  and  $b_2$  is to require that the *asymptotic relative efficiency* (ARE) with respect to the maximum likelihood estimator equals a given value, *e.g.* 95%; the higher the value of ARE, the more the estimate is sensitive to outliers. Marazzi and Ruffieux (1996a) computed optimal values for  $b_1$  and  $b_2$ , depending on the level of contamination and on the nominal level, in considering the underlying distribution to be a Gamma distribution. But less is known about the optimal choice of three tuning parameters, as needed in our algorithm. One could take  $b_1$  and  $b_2$  as shown before and  $b_3$  in the same range. As the target of this paper is not to compute optimal values for the tuning parameter vector  $\mathbf{b} = (b_1, b_2, b_3)$ , we always considered  $b_1 = b_2 = b_3$  in our simulations. This can be justified by the fact that we do not prefer one of the three components of the  $\psi$ -function, *i.e.* we regulate the degree of robustness always in the same way. For example, for the simulation made in the previous section we considered all three tuning constants equal to 1.3.

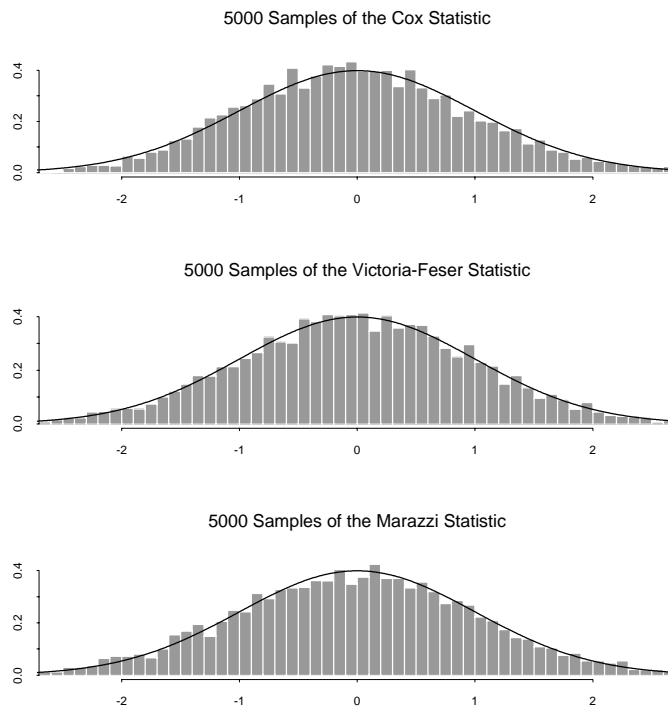
## 2 Simulation Study

### 2.1 Preliminary

In this section we want to answer two questions:

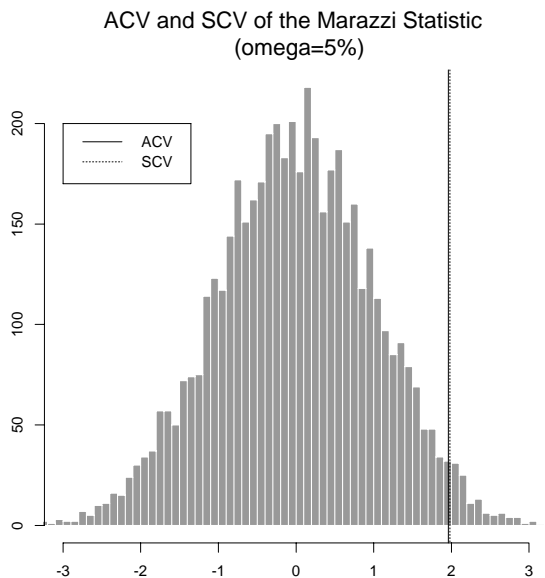
1. Does the asymptotic normality of  $|n^{1/2}\mathbf{U}|$  hold for our simulations? Note that  $\mathbf{U}$  represents one of the three statistics given by equation (1), (2) or (12).
2. Is it reasonable to take as critical value the one of the asymptotic distribution, *i.e.*  $\kappa_{\omega}^*$ , or is it more appropriate to take a simulated critical value?

To do so, we considered as example the hypothesis  $H_0$ : ‘Gamma’ against  $H_1$ : ‘Weibull’. We simulated 5000 samples of size 200 under  $H_0$  and computed the 5000 samples of  $|n^{1/2}\mathbf{U}|$ . The histogram of these samples for each statistic considered is shown by figure 4. The solid lines on the figure represent the density curve of the standard normal distribution. One remarks that the asymptotic normality assertion is justified. In the simulations we considered  $b_1 = b_2 = b_3 = 1.5$  and omitted the presence of contamination.



**Figure 4** Histogram of 5000 samples from each of the three test statistics and the adjusted density curves of the standard normal distribution.

Moreover, as under the asymptotic approach the test rejects  $H_0$  if  $|n^{1/2}\mathbf{U}| > \kappa_\omega^*$ , where  $\kappa_\omega^* = \phi^{-1}(1-\omega/2)$ , we defined  $\kappa_\omega^*$  as the *asymptotic critical value* (ACV). For the simulated approach we defined the *simulated critical value* (SCV) as the empirical quantile of the 5000 samples of desired probability level  $1 - \omega$ . In both cases  $\omega$  represents the nominal level,  $\omega \in \{0.01, 0.03, 0.05, 0.1\}$ . Figure 5 shows an example of the comparison of ACV and SCV. One remarks that there is no significant difference between the two critical values. This statement is also based on the following two facts: We re simulated several times 5000 samples of size 200, and we generated the samples in a first time under  $H_0$  and in a second time under  $H_1$ , *i.e.* from the Weibull distribution. The results are summarised in table 2 for the Gamma case and in table 3 for the Weibull case.



**Figure 5** Histogram of 5000 samples of the Marazzi statistic. The asymptotic, respectively the simulated, critical value under  $H_0$  (Gamma) is represented as a solid, respectively as a dashed line.

$\omega$	Cox		VF		M	
	ACV	SCV	SCV	SCV	SCV	SCV
1%	2.576	2.460	2.594	2.591		
3%	2.170	2.170	2.194	2.250		
5%	1.956	1.929	1.979	1.979		
10%	1.644	1.567	1.671	1.682		

**Table 2** Asymptotic critical value (ACV) and simulated critical value (SCV) under  $H_0$ , *i.e.* generated under Gamma.

$\omega$	Cox		VF		M	
	ACV	SCV	SCV	SCV	SCV	SCV
1%	2.576	2.671	2.526	2.794		
3%	2.170	2.149	2.189	2.263		
5%	1.956	1.963	2.002	2.042		
10%	1.644	1.601	1.618	1.757		

**Table 3** ACV and SCV under  $H_1$ , i.e. generated under Weibull.

Moreover, we calculated the *probability of rejection* (SPoR, respectively APoR) which equals the nominal level in the simulated approach. The results are shown in tables 4 and 5. A further consideration showed that the computed values of APoR are included in the 95%-confidence interval of the nominal level considered. Therefore, we can conclude that there is no significant difference between the asymptotic and the simulated approach. As the use of the asymptotic approach is much faster in time we will consider in our simulations the asymptotic approach.

$\omega$	Cox		VF		M	
	SPoR	AoPR	APoR	APoR	APoR	APoR
1%	1%	0.8%	1.08%	1.02%		
3%	3%	3%	3.16%	3.62%		
5%	5%	4.6%	5.16%	5.36%		
10%	10%	8.4%	10.6%	10.2%		

**Table 4** Simulated probability of rejection (SPoR) and asymptotic probability of rejection (APoR) under  $H_0$ , i.e. generated under Gamma.

$\omega$	Cox		VF		M	
	SPoR	AoPR	APoR	APoR	APoR	APoR
1%	1%	1.26%	0.88%	1.7%		
3%	3%	2.82%	3.08%	3.62%		
5%	5%	5.08%	5.18%	6.14%		
10%	10%	9.24%	9.42%	11.9%		

**Table 5** SPoR and APoR under  $H_1$ , i.e. generated under Weibull.

## 2.2 Simulations

The purpose of robust testing is twofold. First, the level of a test should be stable under small, arbitrary departures from the null hypothesis. Second, the test should still have a good power under small, arbitrary departures from the alternative.

To examine the level and the power of the test, we simulated 1000 samples of  $|n^{1/2}\mathbf{U}|$ , where  $\mathbf{U}$  represents the test statistic, by means of samples of size 200. Moreover, we defined the *attained significance level* by

$$P_{H_0}(H_0 \text{ rejected}) = \frac{\#(|n^{1/2}\mathbf{U}| > \phi^{-1}(1 - \omega/2))}{1000}$$

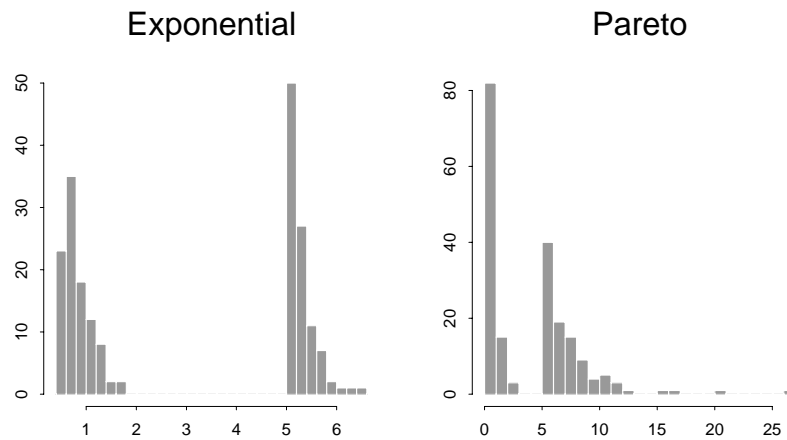
where the samples of size 200 were generated under  $H_0$  and where  $\omega$  represents the nominal level considered. Moreover, we defined the *attained power* by

$$P_{H_1}(H_0 \text{ rejected}) = \frac{\#(|n^{1/2}\mathbf{U}| > \phi^{-1}(1 - \omega/2))}{1000}$$

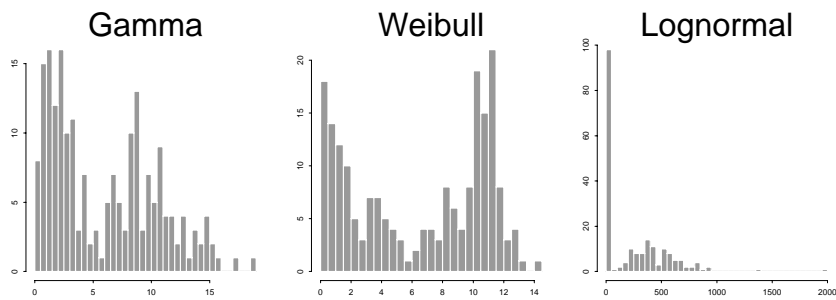
where the samples were generated under  $H_1$ . We have chosen all three tuning parameters equal to 1.5, and for several reasons (see remark 2.2) we did not use the  $\epsilon$ -contamination, but we contaminated our samples in the following way: Let be  $F$  and  $H$  two distributions resulting from the same family (for example Lognormal). If we want to contaminate  $F$  by  $H$ , say with a amount of contamination ( $\epsilon$ ) of 10%, 90% of the sample would contain realisations of the distribution  $F$  and 10% would contain realisations of  $H$ . As we take samples of size 200, the sample would consist in 180 realisations of  $F$  and in 20 realisations of  $H$ . The parameters of  $F$  and  $H$  were chosen in order to obtain a significant difference in looking at the two distributions graphically (see figures 6 and 7). The choices of the parameters for  $F$  and  $H$  are summarised in table 6.

$F$	Parameters for $F$	H	Parameters for $H$	Treated in section(s)
Exponential	$\beta=3, y_0 = 0.5$	Exponential	$\beta=3, y_0 = 5$	2.2.1, 2.2.2
Pareto	$\alpha=3, y_0 = 0.5$	Pareto	$\alpha=3, y_0 = 5$	2.2.1, 2.2.2
Gamma	$\alpha=2, \sigma=1$	Gamma	$\alpha=10, \sigma=1$	2.2.3, 2.2.4
Weibull	$\alpha=1, \sigma=3$	Weibull	$\alpha=9, \sigma=11$	2.2.3
Lognormal	$\theta=e^2, \sigma=e$	Lognormal	$\theta=e, \sigma=e^{1/2}$	2.2.4

**Table 6**  $F$  represents the distribution under which the samples were generated and  $H$  the distribution which was used to contaminate  $F$ .



**Figure 6** Histograms of the contaminated Exponential and Pareto distributions. To show the difference between  $F$  and  $H$  we considered a contamination of 50%.



**Figure 7** Histograms of contaminated Gamma, Weibull and Lognormal distributions. To show the difference between  $F$  and  $H$  we considered a contamination of 50%.

**Remark 2.2:** To know if there is a difference between the contamination using a fixed amount of outliers and the  $\epsilon$ -contamination we simulated the number of outliers in using the  $\epsilon$ -contamination and compared it to the number of outliers in considering our contamination. In simulating their difference 1000 times we obtained the results shown in table 7. We considered 1000 samples of size 200. One remarks that there is no significant difference. This is justified also by the fact that in using the  $\epsilon$ -contamination one gets in average  $200\epsilon$  outliers.

$\epsilon$	Difference
1%	0.0349
3%	0.0600
5%	0.2237
10%	0.1972
50%	0.2016

**Table 7** Comparison of the contamination using a fixed amount of outliers and the  $\epsilon$ -contamination, where  $\epsilon$  represents the amount of outliers. In our simulations we fixed this amount and in using the  $\epsilon$ -contamination one gets in average  $\epsilon$  per cent of outliers. The difference listed in this table represents the average difference (over 1000 samples) between outliers.

Moreover, to obtain the tables presented in the following sections we considered the  $\epsilon$ -contamination as well, but the calculation of the attained significance level and of the attained power did not differ much from the one obtained in using a fixed amount of outliers. Therefore, our choice of contamination is justified, and as the target of this paper is to compare two robust tests it seems to be much nicer to have a fixed amount of outliers. In taking, for example, the  $\epsilon$ -contamination it may be possible that suddenly there a not many outliers. And do not forget that *outliers are candies for robust tests...*

### 2.2.1 Pareto against Exponential

In this section we choose to test the Pareto distribution against the Exponential distribution by means of the three statistics discussed in this paper. The Pareto density we consider is given by

$$f^0(y; \alpha) = \alpha y_0^\alpha y^{-(1+\alpha)} = \frac{\alpha}{y_0} \left( \frac{y}{y_0} \right)^{-(1+\alpha)} \quad (13)$$

where  $0 \leq y_0 \leq y < \infty$ ,  $\alpha$  is the rate parameter ( $\alpha > 0$ ) and  $y_0$  represents the shift. Besides let us remark that if the random variable  $Y$  has a Pareto distribution defined by the density (13) the variable  $Z = \log Y$  has a Exponential density with rate  $\alpha$  and shift  $\log y_0$ .

As alternative hypothesis we consider the Exponential density which is defined as

$$f^1(y; \beta) = \beta \exp(-\beta(y - y_0))$$

where  $y > 0$ ,  $\beta$  is the rate parameter ( $\beta > 0$ ) and  $y_0$  is the shift.

Note that both distributions are particular cases of the Gamma distribution. This results from the fact that an Exponential distribution is a Gamma distribution with shape equal to 1 and with scale equal to the rate parameter. Moreover, we fixed the shift parameter  $y_0$  at 0.5.

One notices that the two distributions are single-parameter distributions, and that the theory presented in chapter 1 concerned double-parameter distributions. In considering only one parameter the equations presented can be simplified, but the main calculations remain the same. Further details and a more explicit presentation of the single-parameter case are given in Victoria-Feser (1996). Note that Victoria-Feser used exactly these two hypotheses to study the robustness properties of the Victoria-Feser statistic given by equation (2) when compared to the Cox statistic given by equation (1).

For the numerical computation of the estimates of the parameters (including the computation of the vector  $\mathbf{c}$  and the matrix  $A$ ) we refer to Marazzi and Ruffieux (1996a) who described an implementation of M-estimators for the Gamma distribution. For the computation of the statistics we used the S-PLUS function `U.for` which is described in the appendix of this paper. In a first time we studied the attained significance level in simulating Pareto samples which were computed by means of the S-PLUS function `rpseudo` (see appendix). And in a second time we took a look at the attained power in generating Exponential samples. The results of the simulations are summarised in tables 8 and 9: Table 8 represents the attained significance level (in %) and table 9 represents the attained power. We refer to table 6 for further details on the choice of the parameters and on the contamination chosen.

In considering table 8 we observe that the original Cox statistic has a very strange behaviour with small amounts of contamination, and with an amount of contamination of

10% the null hypothesis has always been rejected. On the other hand, we observe for the Marazzi statistic that small departures from the model do not influence the level of the test at least for amounts up to 5%. This is not the case with the Victoria-Feser statistic which over rejects  $H_0$ , *i.e.* the Pareto distribution, already at  $\epsilon = 3\%$ . Hence, we conclude that the Marazzi statistic is more stable. Moreover, the fact that the three statistics do not reach the nominal level exactly is due to the fact that we used the asymptotic critical value instead of the simulated critical value (see section 2.1).

Nominal levels (in %)												
Amount of contamination	1%			3%			5%			10%		
	Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
0%	1.6	0.9	1.6	3.1	3.1	4.1	4.7	6.5	5.8	8.9	9.8	14
3%	73.1	4.2	1.3	83.6	7.9	5	89.3	13.4	7.6	95.8	20.8	13.8
5%	74.5	12.9	3	87.7	29.9	7.7	91	37	10.5	95.9	47.7	18.4
10%	100	96.2	16.2	100	99.6	26.6	100	99.8	14	100	99.9	51.3

**Table 8** Attained significance levels (in %) in using the original Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) in using a fixed amount of contamination (Pareto against Exponential).

Furthermore, table 9 indicates that the Marazzi statistic is the most powerful statistic considered. One also remarks that the Cox statistic did never reject  $H_0$  even when the samples were generated under  $H_1$ , *i.e.* from an Exponential distribution. A study of this lack of power of the Cox statistic is given in remark 2.2.1.

Nominal levels (in %)												
Amount of contamination	1%			3%			5%			10%		
	Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
0%	0	0.806	0.787	0	0.912	0.879	0	0.945	0.911	0	0.972	0.949
3%	0	0.67	0.626	0	0.799	0.771	0	0.844	0.839	0	0.912	0.89
5%	0	0.356	0.49	0	0.543	0.671	0	0.654	0.695	0	0.764	0.826
10%	0	0	0.103	0	0.002	0.218	0	0.007	0.3	0	0.002	0.424

**Table 9** Attained power in using the original Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) with fixed contamination (Pareto against Exponential).

**Remark 2.2.1:** To study the lack of power of the Cox statistic we computed percentiles of the Pareto distribution by means of `ppareto` (see appendix). Once computed these percentiles we calculated the Cox statistic in considering the test  $H_0$ : Pareto against  $H_1$ : Exponential. The aim of this simulation was to know how many percentiles are needed to obtain a test who reaches the nominal level exactly. The results are summarised in table 10. The decision was based on the comparison of the observed statistic to the asymptotic critical value at the nominal level 5%, *i.e.* to 1.96. One notices that more than 10000 percentiles are needed to obtain the result desired, *i.e.* the acceptance of  $H_0$ . A further consideration showed that the size of the sample considered has to be larger than 16700! A sample of size 16700 is not realistic for a real data set, hence one should be very careful in using the classical Cox statistic.

Number of percentiles	$U_{cox}$	Decision
200	20.96	$H_1$
500	13.03	$H_1$
1000	9.06	$H_1$
5000	3.83	$H_1$
10000	2.61	$H_1$
20000	1.76	$H_0$
50000	1.02	$H_0$

**Table 10** *Simulated Pareto percentiles, the corresponding value of the Cox statistic and the decision taken (Pareto against Exponential).*

In generating under the same hypotheses, *i.e.*  $H_0$ : Pareto against  $H_1$ : Exponential, Exponential percentiles we remarked that a sample of size 470 is needed to obtain the rejection of  $H_0$ . The results are summarised in table 11. Therefore, we can conclude that the Cox statistic is not able to differ between the Pareto and the Exponential statistic for small samples.

Number of percentiles	$U_{cox}$	Decision
200	1.08	$H_0$
470	1.96	$H_1$
500	2.02	$H_1$
1000	3.00	$H_1$
10000	9.92	$H_1$

**Table 11** *Simulated Exponential percentiles, the observed Cox statistic and the decision taken (Pareto against Exponential).*

### 2.2.2 Exponential against Pareto

In this section let the null hypothesis  $H_0$  be that the probability density distribution function is Exponential, namely

$$f^0(y; \beta) = \beta \exp(-\beta(y - y_0))$$

where  $y > 0$ , and let  $H_1$  be the hypothesis that the probability density function is Pareto, namely

$$f^1(y; \alpha) = \alpha y_0^\alpha y^{-(1+\alpha)}$$

where  $0 \leq y_0 \leq y < \infty$ . For further details on the computation we refer to section 2.2.1. The results of our simulations are summarised in tables 12 and 13.

Table 12 represents the attained significance levels in considering several amounts of contamination and several values of the nominal level. One remarks that the Cox statistic always rejects  $H_0$  for contaminated samples. This result is not surprising as the breakdown point, *i.e.* the highest percentage of outliers that will not completely distort the estimate, of the maximum likelihood estimate  $\hat{\beta}$  equals 0, and as with an amount of contamination of 3% we already have 6 outliers in our sample of size 200. Moreover, one notices that even the Victoria-Feser statistic, which uses the maximum likelihood estimate as well, over rejects  $H_0$  at  $\epsilon = 3\%$ . This is not the case with the Marazzi statistic which remains stable at amounts of outliers up to 5%.

Nominal levels (in %)												
Amount of contamination	1%			3%			5%			10%		
	Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
0%	0.4	0.4	1	1.7	2.2	2.6	3.4	5.4	5.7	8.1	8.7	11.6
3%	100	10.3	1.2	100	21.3	4.9	100	28.8	8.8	100	40	14.8
5%	100	66.6	3.2	100	85.2	5.9	100	89.5	10	100	95.3	18.3
10%	100	100	15.7	100	100	25.1	100	100	34.8	100	100	49.4

**Table 12** Attained significance levels (in %) in using the original Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) in considering a fixed amount of contamination (Exponential against Pareto).

Table 13 represents the attained powers. One observes that, once again, the Marazzi statistic is the most powerful.

Amount of contamination	Nominal levels (in %)											
	1%			3%			5%			10%		
	Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
0%	0.293	0.561	0.817	0.332	0.714	0.907	0.348	0.79	1	0.379	0.827	1
3%	0.234	0.995	1	0.264	0.995	1	0.248	0.999	1	0.255	0.999	1
5%	0.221	0.992	1	0.264	0.996	1	0.274	0.999	1	0.308	0.998	1
10%	0.243	0.956	1	0.277	0.984	1	0.277	0.989	1	0.292	0.995	1

**Table 13** Attained power in using the original Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) with fixed contamination (Exponential against Pareto).

**Remark 2.2.2:** Based on remark 2.2.1 we also simulated Pareto percentiles to know the size of the sample needed to obtain the non-significance of  $H_0$  in using the Cox statistic (see table 14). One notices that the Cox statistic works very well, even for small samples.

Number of percentiles	$U_{cox}$	Decision
200	22.04	$H_1$
500	16.13	$H_1$
1000	16.29	$H_1$
5000	23.30	$H_1$
10000	31.55	$H_1$
20000	43.87	$H_1$
25000	48.84	$H_1$
50000	68.58	$H_1$
100000	96.63	$H_1$

**Table 14** Simulated Pareto percentiles, the corresponding value of the Cox statistic and the decision taken (Exponential against Pareto).

### 2.2.3 Gamma against Weibull

As null hypothesis we consider the Gamma distribution defined by its density

$$f^0(y; (\alpha, \sigma)) = \frac{1}{\sigma \Gamma(\alpha)} \left(\frac{y}{\sigma}\right)^{\alpha-1} \exp(-(y/\sigma)) \quad (14)$$

where  $y > 0$ ,  $\sigma > 0$ ,  $\alpha > 0$  and  $\Gamma$  is the Gamma function. The pair  $(\sigma, \alpha)$  is the usual parameter vector;  $\alpha$  is the shape parameter, and  $\sigma$  is the scale parameter. And as alternative hypothesis we consider the Weibull distribution with density

$$f^1(y; (\alpha, \sigma)) = \frac{\alpha}{\sigma} \left(\frac{y}{\sigma}\right)^{\alpha-1} \exp(-(y/\sigma)^\alpha)$$

where  $y > 0$ ,  $\sigma > 0$  and  $\alpha > 0$ .

The case of the Gamma distribution is studied by Marazzi and Ruffieux (1996a), and the case of the Weibull distribution is discussed in Marazzi (1997c) who adapted the implementation of M-estimators for the Gamma distribution described in Marazzi and Ruffieux (1996a) to the Weibull distribution. The results of our simulations are summarised in tables 15 and 16.

Nominal levels (in %)												
Amount of contamination	1%			3%			5%			10%		
	Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
0%	0.5	0.5	1.5	2.6	2.6	4	4.4	4.6	5.8	9.3	10.1	10.5
3%	71.6	3.8	2.4	83.6	8	5.9	86.3	11.3	7.8	93.4	19.6	16
5%	92.4	13.7	3.6	96.4	26.2	10	97.3	37.6	13.9	99.2	50.7	22.4
10%	98.5	86.5	31.5	99	94.5	48.3	99.6	96.3	56.9	99.9	98.7	67.7

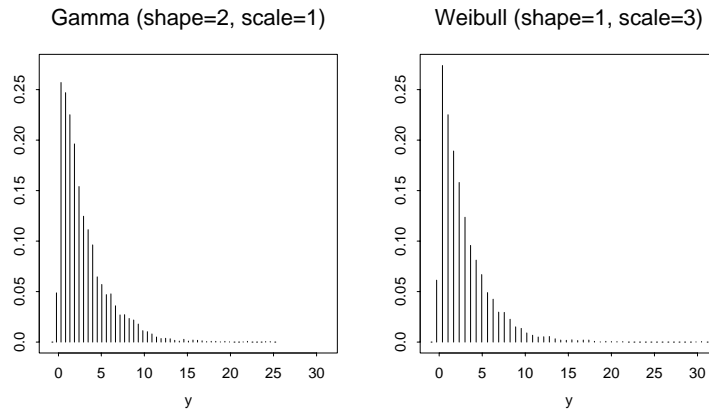
**Table 15** Attained significance levels (in %) in using the Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) with a fixed amount of contamination (Gamma against Weibull).

In considering table 15 we observe that the Marazzi statistic is very stable for amounts up to 5%. For the Victoria-Feser statistic we observe a significant over rejection of  $H_0$  at 3% yet. The Cox statistic seems to be the less appropriated statistic when contamination is considered. Furthermore, in table 16 one observes that in taking the Cox statistic the power decreases if the amount of outliers increases. This seems to be logical as the statistic is very sensible to outliers. In using a robust version of the Cox statistic one observes the opposite, *i.e.* the power increases together with the amount of contamination. This fact is

very satisfying. Moreover, one remarks that the Marazzi statistic is more powerful than the Victoria-Feser statistic. The small power of the tests, even with no contamination, results in the fact shown by figure 8, where a histogram for the Gamma distribution, respectively for the Weibull distribution, is drawn. One remarks that there is no graphical difference between the two distributions, hence all statistics encounter problems in testing between the two distributions. But, this example shows clearly that even under these circumstances the Marazzi statistic reaches a higher power compared to the Cox and the Victoria-Feser statistic.

Nominal levels (in %)												
Amount of contamination	1%			3%			5%			10%		
	Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
0%	0.009	0.013	0.019	0.038	0.032	0.053	0.058	0.051	0.085	0.104	0.084	0.103
3%	0.008	0.021	0.016	0.021	0.053	0.051	0.028	0.073	0.072	0.084	0.15	0.149
5%	0.005	0.035	0.035	0.018	0.089	0.071	0.02	0.128	0.13	0.059	0.214	0.181
10%	0.002	0.112	0.141	0.012	0.236	0.272	0.029	0.271	0.332	0.053	0.391	0.459

**Table 16** Attained power in using the original Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) with contamination (Gamma against Weibull).



**Figure 8** Histogram of the Gamma distribution, respectively Weibull distribution, in using the parameters described in table 6.

### 2.2.4 Gamma against Lognormal

As null hypothesis we consider the Gamma density which is defined by equation (14) in section 2.2.3. As alternative hypothesis we consider the Lognormal density which is defined as

$$f^1(y; (\theta, \sigma)) = \frac{1}{\sigma y \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{\log y - \theta}{\sigma}\right)^2\right)$$

where  $y > 0$  and  $\sigma > 0$ .

The attained significance levels are summarised in table 17. The Marazzi statistic is stable up to an amount of 3%, and the Victoria-Feser statistic and the Cox statistic are not stable at all. For example, at a nominal level of 5% and an amount of 3% the attained significance level for the Victoria-Feser statistic equals 14%, whereas the one of the Marazzi statistic remains stable.

		Nominal levels (in %)											
		1%			3%			5%			10%		
Amount of contamination		Cox	VF	M	Cox	VF	M	Cox	VF	M	Cox	VF	M
	0%		0.8	0.8	1.5	2.5	3.1	3.5	4	4.4	5.3	7.5	10.3
3%		24.9	3.4	1.9	44.9	9.2	5.3	53.6	14	6.7	67.4	19.9	13.7
5%		52.3	15.4	3.7	72.3	25.9	10.4	79.9	36.1	13.9	88.9	49.7	25.2
10%		83.3	84.5	30.9	92.8	92.1	49.2	96.8	94.7	56.1	98.1	97.9	68.8

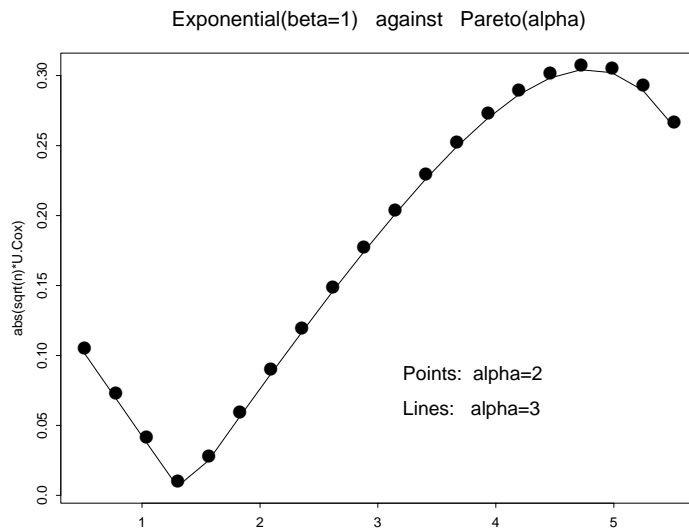
**Table 17** Attained significance levels (in %) in using the original Cox statistic (Cox), the Victoria-Feser statistic (VF) and the Marazzi statistic (M) with contamination (Gamma against Lognormal).

### 3 Outstanding Questions

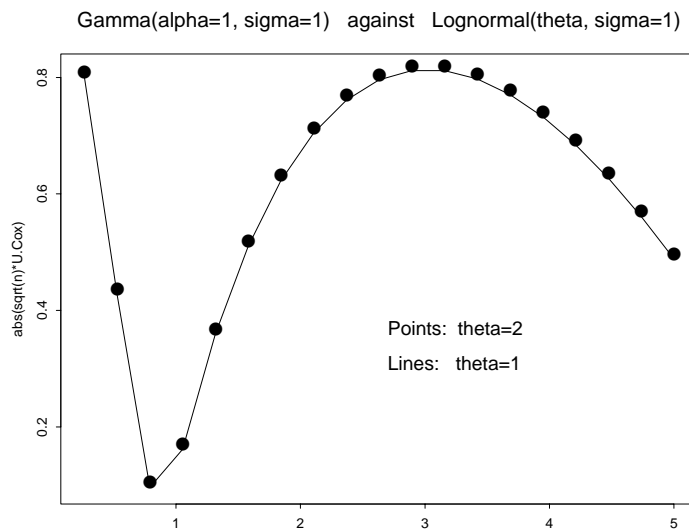
In the simulation study presented in the previous section we remarked two problems in using the original Cox statistic:

1. In section 2.2.1, where we tested the Pareto against the Exponential distribution, we wanted to explain the lack of power who occurred when using the original Cox statistic proposed by Cox (1961, 1962). We noticed that to compute a Cox statistic who would be exact a sample of size 17400 is required (see remark 2.2.1). Therefore we concluded that the Cox statistic is **not able to differ between the Pareto and the Exponential distribution** in considering small samples. Note that small data sets represent the bulk of real data sets. We did not understand the reason(s) why this phenomena occurred. Hence a further consideration should be undertaken.
2. We realised another strange behaviour in using the Cox statistic which frightens us more than the first one. We noticed that under some circumstances the **statistic does not depend on the parameters**. An example of such a circumstance is shown by figure 9. Suppose that we want to test  $H_0$ : Exponential with rate  $\beta$  equal to 1 against  $H_1$ : Pareto with rate  $\alpha$  - the shift parameter  $y_0$  is fixed at 0.5. As sample we considered a sequence of 20 evenly spaced points from 0 to 5. In computing the Cox statistic for each point of this sequence in considering in a first time  $\alpha = 2$  and in a second time  $\alpha = 3$  we remarked that the computed values of the Cox statistic did not change. Moreover, we noticed that the expectation of the difference of the log likelihood functions for each of the two hypothesis remained constant (the constant equalled -0.685). You may think that this phenomena may be caused by the violation of the hypothesis that  $H_0$  and  $H_1$  should be separate in the sense that an arbitrary simple hypothesis in  $H_0$  cannot be obtained as a limit of simple hypotheses in  $H_1$  (Cox, 1961). But this does not seem to be the case. Indeed, we noticed the same phenomena in considering the test  $H_0$ : Gamma with shape  $\alpha$  and scale  $\sigma$  equal to 1 against  $H_1$ : Lognormal with varying mean  $\theta$  and  $\sigma$  equal to 1, where for  $\theta$  we choose 2 and 1. The result is shown by figure 10. In this case the hypothesis of having separate hypotheses  $H_0$  and  $H_1$  is certainly verified. Once again, we did not understand the reason(s) why this phenomena occurred.

It may be interesting to study in a further work these two problems. We did not investigate further time in answering these two outstanding questions, as our main target was to compare robust statistics. However, a more detailed presentation, especially for the computations made, are available from the author.



**Figure 9** The calculated Cox statistics of 20 evenly spaced points for the test between  $H_0$ : Exponential and  $H_1$ : Pareto. The parameter of the Exponential model was fixed at 1 ( $\beta$ ), and for the one of the Pareto model, i.e.  $\alpha$ , we choose 2 ( $\rightsquigarrow$  points) and 3 ( $\rightsquigarrow$  lines). One remarks that for both cases the computed Cox statistics are the same.



**Figure 10** The calculated Cox statistics of 20 evenly spaced points for the test between  $H_0$ : Gamma and  $H_1$ : Lognormal. The parameters of the Gamma model were fixed at 1, and the scale parameter  $\sigma$  of the Lognormal distribution was fixed at 1. For the mean parameter of the Lognormal, i.e.  $\theta$ , we choose 2 ( $\rightsquigarrow$  points) and 1 ( $\rightsquigarrow$  lines). Hence, the Cox statistic does not depend on the parameter  $\theta$ .

## 4 Summary and Conclusions

In this paper three statistics to test between two income models were proposed:

1. The **Cox statistic** which is a modification of the Neyman-Pearson maximum likelihood ratio test. This test is initialised by taking the classical maximum likelihood estimates of the parameters (Cox, 1961 and 1962). The statistic is given by equation (1).
2. The **Victoria-Feser statistic** which robustly bounds the test statistic proposed by Cox. The initialisation is done by taking the maximum likelihood estimates of the parameters (Victoria-Feser, 1996). The computation algorithm is presented in section 1.1, and the test statistic is given by equation (2).
3. The **Marazzi statistic** which uses the idea of bounding the test statistic proposed by Victoria-Feser. But instead of taking the maximum likelihood estimates, the  $B_g^p$ -estimates proposed by Hampel (1986) are taken. The computation algorithm of this statistic is given in section 1.2, whereas equation (12) represents the final form of the test statistic.

To compare these three statistics we considered several simulated examples. The common features observed can be summarised as follows:

- In general, the Marazzi statistic is stable up to an amount of 3% to 5% of contamination. In this context it is important to say that we always used a fixed amount of contamination. Moreover, the Victoria-Feser statistic is compared to the Marazzi statistic less stable. This results from the fact that Victoria-Feser considers the maximum likelihood estimates of the parameters which have a breakdown point of 0. And finally, we remarked that the Cox statistic is not stable at all.
- Our simulations showed that the Marazzi statistic is the most powerful statistic considered. In the case where we tested the Gamma distribution against the Weibull distribution (see section 2.2.3) we even remarked that the power of the Marazzi statistic increases together with the amount of contamination.

Therefore in general, we would recommend the Marazzi statistic because of its superior overall performance when contamination is present; this is usually the case with real data sets.

But there are still some unresolved problems:

- A. The strange behaviour of the Cox statistic presented in section 3.
- B. The optimal values for the tuning parameters  $b_1$ ,  $b_2$  and  $b_3$  which are used to compute the Marazzi statistic. We did not take in this problem and therefore we considered  $b_1 = b_2 = b_3$  in our simulations to regulate the degree of robustness always in the same way.
- C. The problem of analytically computing the breakdown point of the  $B_s^p$ -estimators for the Gamma and the Weibull models is still unsolved, although some exploratory numerical computations suggest that the contamination bias of these estimators is reasonably low for practical problems.

A further work would consist in solving these outstanding problems. Furthermore, to confirm the conclusions achieved more examples should be treated. We are sure that these problems will be resolved in the near future. To contribute, we included the procedures and functions described in this paper into S-PLUS libraries (for details see appendix). They are made available under <http://www.hospvd.ch/iumpsp/download/menu.htm>.

*'Uncertainty ... Something You can  
Always Count On'*

## Acknowledgement

The author is grateful to Alex Randriamiharisoa for his generous help with programming. He made the simulation study feasible. Moreover, the author wants to thank Prof. Alfio Marazzi for his helpful comments and suggestions. A special thank goes to M. Genton and E. Chavez of the chair of applied statistics at the Swiss Federal Institute of Technology.

This research was presented as a semester project to the faculty of the Department of Mathematics at the Swiss Federal Institute of Technology in Lausanne.

## A Appendix

### A.1 S-Plus Functions for the Pareto Distribution

The Pareto density we consider is given by

$$f(y; \alpha) = \alpha y_0^\alpha y^{-(1+\alpha)}$$

where  $0 \leq y_0 \leq y < \infty$ ,  $\alpha$  is the rate parameter ( $\alpha > 0$ ) and  $y_0$  represents the shift. Hence, the point distribution function is

$$F(y; \alpha) = 1 - \left(\frac{y_0}{y}\right)^\alpha$$

A natural way to simulate a distribution is to consider its inverse and to use uniform deviates. Indeed, let  $F$  be the distribution of the random variable  $Y$  and let  $U$  be a variable which distribution function is  $U(0, 1)$ , then the distribution of  $Z = G^{-1}(U)$  is

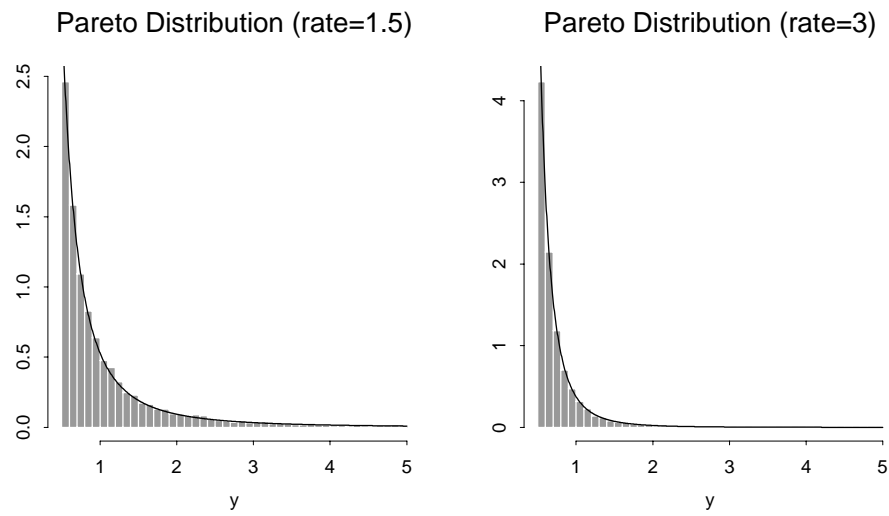
$$\begin{aligned} P(Z \leq z) &= P(G^{-1}(U) \leq z) \\ &= P(U \leq G(z)) \\ &= G(z) \end{aligned}$$

In other words, one remarks that  $Y$  and  $Z$  have the same distribution function. This is exactly what the S-PLUS function `rpareto` does:

```
rpareto <- function(size, alpha=3, y0=0.5)
{
  u <- runif(size, min = 0, max = 1)
  z <- y0/(1-u)^(1/alpha)
  z
}
```

A visual verification is shown by figure 11, where the solid line represents the exact density function. In the same way the S-PLUS function `ppareto` generates percentiles of the Pareto distribution.

```
ppareto <- function(size, alpha=3, y0=0.5)
{
  u <- seq(from = 1e-08, to = 1-1e-08, length = size)
  z <- y0/(1-u)^(1/alpha)
  z
}
```



**Figure 11** Histograms of two Pareto samples generated by means of the S-PLUS function `rpareto`. The histograms were truncated at  $y = 5$ . The solid line represents the exact density function.

## A.2 S-Plus Functions for the Computation of the Statistics

The functions used in this paper have been developed using parts of the subroutine library ROBETH described in Marazzi (1993). Several choices concerning the specification of the estimators and the specification of the computational algorithms have been made, hence the user cannot change many of these choices. The programs for the computation of the M-estimators of the parameters of the Gamma distribution (Marazzi, 1997b), of the Weibull distribution (Marazzi, 1997c) and of the Lognormal distribution (Marazzi, 1996b) are available as subsets of the library ROBETH. To guarantee their interface to S-PLUS all functions are available in form of the following S-PLUS libraries: `robeth`, `robgam`, `robwbl` and `robrnm`.

Moreover, to access the functions used in this paper the library `dcompar` is made available. This library contains the functions which purposes are:

<code>U.for</code>	Computation of the Cox, the Victoria-Feser or the Marazzi statistic.
<code>M.L.exp</code>	Maximum likelihood estimate $\hat{\beta}$ for the Exponential distribution.
<code>M.L.pareto</code>	Maximum likelihood estimate $\hat{\alpha}$ for the Pareto distribution.
<code>M.E.exp</code>	M-estimate $\hat{\beta}$ for the Exponential distribution.
<code>M.E.pareto</code>	M-estimate $\hat{\alpha}$ for the Pareto distribution.
<code>AV.ML.exp</code>	Asymptotic covariance matrix of the maximum likelihood estimate $\hat{\beta}$ .
<code>AV.ML.pareto</code>	Asymptotic covariance matrix of the maximum likelihood estimate $\hat{\alpha}$ .
<code>AV.ME.exp</code>	Asymptotic covariance matrix of the M-estimate $\hat{\beta}$ .
<code>AV.ME.pareto</code>	Asymptotic covariance matrix of the M-estimate $\hat{\alpha}$ .
<code>Tab.exp</code>	Computation and storage of $(a_{11}, a_{21}, a_{22})$ and $(c_1, c_2)$ .
<code>Tab.pareto</code>	Computation and storage of $(a_{11}, a_{21}, a_{22})$ and $(c_1, c_2)$ .

The main function of this library is `U.for` which algorithms are programmed in FORTRAN using parts of the subroutine library `robeth`. The description of this function is given later on.

For further details on the functions concerning the Pareto and the Exponential distribution we refer to Marazzi (1996a and 1997b) as both distributions are particular distributions of the Gamma family.

---

**function U.for**

Computation of the Cox, the Victoria-Feser or the Marazzi statistic for choosing between two asymmetric distribution models.

---

**Specification**

```
function U.for(y, Tab0, Tab1, b3,
              c3alg="N", f0p1=NA, f0p2=NA, f1p1=NA, f1p2=NA,
              maxta=2, maxtc=2, maxit=30, monit=1, tol=0.0001,
              y0=0)
```

**Purpose**

This function computes the Cox statistic given by equation (1), the Victoria-Feser statistic given by equation (2) or the Marazzi statistic given by equation (12).  $(a_{11}, a_{21}, a_{22})$  and  $(c_1, c_2)$  are assumed to be given as input in form of `Tab0` for the null hypothesis and in form of `Tab1` for the alternative hypothesis. They must be computed before calling `U.for` by means of the functions described in Marazzi (1996a, 1997b and 1997c) for the Lognormal, Gamma and Weibull cases. For the Pareto and the Exponential cases the functions `Tab.pareto` and `Tab.exp` should be taken. The following table summarises the arguments that must be used to compute a particular statistic:

Statistic	Tab0 and Tab1	b3	f0p1 and f0p2	f1p1 and f1p2
Cox	Computed by means of $b_1 = b_2 = 300$ .	300	ML-estimates of the parameters.	ML-estimates of the parameters.
Victoria-Feser	Computed by means of $b_1 = b_2 = 300$ .	1.5	ML-estimates of the parameters.	ML-estimates of the parameters.
Marazzi	Computed by means of $b_1 = b_2 = 1.5$ .	1.5	M-estimates of the parameters (per default).	M-estimates of the parameters (per default).

The user has the possibility to choose other values for the tuning parameters.

## Arguments

<code>y</code>	Observation vector.
<code>Tab0</code>	<code>Tab0</code> contains $(c_1, c_2, a_{11}, a_{21}, a_{22})$ in considering the distribution assumed in the null hypothesis. This values have to be computed before calling <code>U.for</code> (see purpose).
<code>Tab1</code>	<code>Tab0</code> contains $(c_1, c_2, a_{11}, a_{21}, a_{22})$ for the distribution considered in the alternative hypothesis.
<code>b3</code>	Tuning parameter $b_3$ used to bound the statistic.
<code>c3alg</code>	Per default, <i>i.e.</i> <code>c3alg="N"</code> , the Newton algorithm is used for the computation of the solution $c_3$ of the integral given in equation (10). If <code>c3alg</code> is set to something else the Regula Falsi algorithm is used.
<code>f0p1</code> <code>f0p2</code> <code>f1p1</code> <code>f1p2</code>	Per default, the M-estimates of the parameters are computed. The maximum likelihood estimates have to be calculated before calling <code>U.for</code> . <code>f0p1</code> represents the estimate of the main parameter of the distribution considered in the null hypothesis, and <code>f0p2</code> the estimate of the nuisance parameter. <code>f1p1</code> and <code>f1p2</code> are the estimates of the parameters in considering the alternative distribution. The following table summarises the required arguments in considering the maximum likelihood estimates:

	<code>f.p1</code>	<code>f.p2</code>	
Exponential	1	$\hat{\beta}$	<code>y0= 0.5</code>
Pareto	1	$\hat{\alpha}$	<code>y0= 0.5</code>
Gamma	$\hat{\alpha}$	$\hat{\sigma}$	
Weibull	$\hat{\alpha}$	$\hat{\sigma}$	
Lognormal	$\hat{\theta}$	$\hat{\sigma}$	

<code>maxta</code>	Maximum number of steps for improving $(a_{31}, a_{32}, a_{33})$ (see equation (11)).
<code>maxtc</code>	Maximum number of steps for improving $c_3$ (see equation (10)).
<code>maxit</code>	Maximum number of cycles of the main algorithm.
<code>monit</code>	If <code>monit</code> $>$ 0 and the iteration counter is a multiple of <code>monit</code> , the current values of the equations (10) and (11) are displayed. If no iteration monitoring is required, set <code>monit</code> equal to 0 or to a negative integer value.
<code>tol</code>	Required relative precision of $(a_{31}, a_{32}, a_{33})$ . As the convergence of $c_3$ is much faster, the required relative precision of $c_3$ will be smaller than <code>tol</code> . Therefore, it is sufficient to require a relative precision for $(a_{31}, a_{32}, a_{33})$ .
<code>y0</code>	The shift parameter $y_0$ , which is used for testing Pareto against Exponential and vice versa.

**Value**

List with the following components

<b>U</b>	The test statistic $U$ multiplied by $n^{1/2}$ .
<b>nit</b>	Reached number of iterations.
<b>bug</b>	This variable equals 1 if there was an error while proceeding the computation. Indeed, in this case the inverse of the matrix $A$ could not be calculated. Note that in this situation the <b>U</b> value is set to 0, the expectation of the asymptotic distribution. If <b>bug</b> = 0, everything worked well.
<b>a3</b>	Vector containing $a_{31}$ , $a_{32}$ and $a_{33}$ given as solutions of equations (11).
<b>c3</b>	Component containing $c_3$ given as solution of equation (10).
<b>fa3</b>	The values of the three integrals given in equations (11) in substituting $a_{31}$ , $a_{32}$ and $a_{33}$ . They should be all equal to 0 if <b>tol</b> has been set sufficiently small. Note that for $a_{33}$ the value of the integral minus 1 is printed.
<b>fc3</b>	The values of equation (10) in substituting $c_3$ .
<b>fpar</b>	The estimates for the parameters of the two distributions.
<b>ac</b>	The used values of <b>Tab0</b> , <i>i.e.</i> $(a_{11}, a_{21}, a_{22}, c_1, c_2)$ .
<b>ExpL</b>	The expectation of the difference of the log likelihood functions for each of the two hypotheses.
<b>call</b>	The calling sequence.

**Example**

Suppose that we want to test the Weibull distribution against the Gamma distribution in considering a sample **sam** of size 200 generated by means of a Weibull distribution with  $\alpha = 1$  and  $\sigma = 3$ :

```
> sam <- rweibull(200, 1, 3)
```

To use the functions described before, type:

```
> library(robeth, T); library(robgam, T); library(robwbl, T)
```

```
> library(dcompar, T)
```

First, we can calculate the classical Cox statistic given by equation (1) by means of:

```
> Tabw.ml <- Tab.weibul(b1=300,b2=300)
> Tabg.ml <- Tab.gamma(b1=100,b2=100,t1l=0.0001,tol=0.0001)
> ew <- M.L.weibul(sam)
> eg <- M.L.gamma(sam)
```

where `ew` and `eg` contain the classical maximum likelihood estimates of the parameters.

```
> u.cox <- U.for(sam, Tab0=Tabw.ml, Tab1=Tabg.ml, b3=300,
                 f0p1=ew$alpha, f0p2=ew$sigma, f1p1=eg$alpha ,f1p2=eg$sigma)
```

The value `|u.cox$U|` is smaller than  $\phi^{-1}(1-\omega/2)$ , where for example  $\omega = 5\%$ , hence there is no evidence against the Weibull distribution. Moreover, to calculate the Victoria-Feser statistic given by equation (2), type:

```
> u.vf <- U.for(sam, Tab0=Tabw.ml, Tab1=Tabg.ml, b3=1.5,
                f0p1=ew$alpha, f0p2=ew$sigma, f1p1=eg$alpha, f1p2=eg$sigma)
```

Once again there is no evidence for the Gamma distribution. And finally, the Marazzi statistic given by equation (12) can be computed as follows:

```
> Tabw.m <- Tab.weibul(b1=1.5,b2=1.5)
> Tabg.m <- Tab.gamma(b1=1.5,b2=1.5)
> u.m <- U.for(sam, Tab0=Tabw.m, Tab1=Tabg.m, b3=1.5)
```

## References

- ATKINSON A.C. (1970) *A Method for Discriminating between Models*. Journal of the Royal Statistical Society, Serie B 32, 323-353.
- COX D.R. (1961) *Tests of Separate Families of Hypotheses*. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, Berkeley, 105-123. University of California Press.
- COX D.R. (1962) *Further Results on Tests of Separate Families of Hypotheses*. Journal of the Royal Statistical Society, Serie B 24, 406-424.
- HAMPEL F.R., RONCHETTI E.M., ROUSSEEUW P.J. and STAHEL W.J. (1986) *Robust Statistics, the Approach Based on Influence Functions*. Wiley, New York.
- HERITIER S. and RONCHETTI E.M. (1994) *Robust Bounded-Influence Tests in General Parametric Models*. Journal of the American Statistical Association 89(427), 897-904.
- MARAZZI A. (1993) *Algorithms, Routines and S-functions for Robust Statistics*. Wadsworth & Brooks Cole, Belmont, California, Reprinted by Chapman and Hall, New York.
- MARAZZI A. and RUFFIEUX C. (1996a) Implementing M-estimators of the Gamma Distribution. In: H. Rieder (Ed.), *Robust Statistics, Data Analysis, and Computer Intensive Methods, In Honor of Peter Huber's 60th Birthday*. Lecture Notes in Statistics, 109, Springer Verlag, Heidelberg.
- MARAZZI A. and RANDRIAMIHARISOA A. (1996b) *Implementing Robust Estimators of the Gauss Distribution with S-plus Functions*. Draft Copy. Institut universitaire de medecine sociale et preventive, Bugnon 17, CH-1005 Lausanne.
- MARAZZI A. (1997a) *Estimating and Testing the Means of Asymmetric Distributions using M-estimators*. Technical Report. Institut universitaire de medecine sociale et preventive, Bugnon 17, CH-1005 Lausanne.
- MARAZZI A. and RANDRIAMIHARISOA A. (1997b) *S-plus Functions for M-estimators of the Parameters of the Gamma Distribution*. Technical Report. Institut universitaire de medecine sociale et preventive, Bugnon 17, CH-1005 Lausanne.
- MARAZZI A. and RANDRIAMIHARISOA A. (1997c) *S-plus Functions for M-estimators of the Parameters of the Weibull Distribution*. Technical Report. Institut universitaire de medecine sociale et preventive, Bugnon 17, CH-1005 Lausanne.
- VICTORIA-FESER M.-P. (1996) *Choosing between two Income Distribution Models with Contaminated Data*. Discussion Paper No. DARP 18, London School of Economics, London WC2A 2AE.