Working Paper Series, N. 2, March 2010

An overview of robust methods in medical research

Alessio Farcomeni
Department of Hygene and Public Health
Sapienza University of Rome
Italy

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: Robust statistics is an extension of classical parametric statistics that specifically takes into account the fact that the assumed parametric models used by the researchers are only approximate. In this paper we review and outline how robust inferential procedures may routinely be applied in practice in the biomediacal research. Numerical illustrations are given for the t-test, regression models, logistic regression, survival analysis and ROC curves, showing that robust methods are often more appropriate than standard procedures.

Keywords: Breakdown point, Influence function, Likelihood methods, Logistic regression, M-estimation, Regression-scale model, `R`, ROC curve, Student t-test, Survival analysis.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Laura Ventura
tel: +39 049 827 4177
ventura@stat.unipd.it
http://www.stat.unipd.it/~ventura

An overview of robust methods in medical research

Alessio Farcomeni
Department of Hygene and Public Health
Sapienza University of Rome
Italy

Laura Ventura
Department of Statistical Sciences
University of Padua
Italy

Abstract: Robust statistics is an extension of classical parametric statistics that specifically takes into account the fact that the assumed parametric models used by the researchers are only approximate. In this paper we review and outline how robust inferential procedures may routinely be applied in practice in the biomediacal research. Numerical illustrations are given for the t-test, regression models, logistic regression, survival analysis and ROC curves, showing that robust methods are often more appropriate than standard procedures.

Keywords: Breakdown point, Influence function, Likelihood methods, Logistic regression, M-estimation, Regression-scale model, `R`, ROC curve, Student t-test, Survival analysis.

1 Introduction

The normal distribution is the basis of statistical analyses in medicine, genetics and in related sciences. Indeed, standard estimation and testing procedures are often carried out based on the normality assumption. More precisely, parametric inferential procedures based on the sample mean, the standard deviation, the one- and two-samples t-test, and so on, are the most efficient under this assumption. However, it is well-known that they are not robust when the normal distribution is just an approximate parametric model or in the presence of deviant values in the observed data, since they can be upset completely by a single outlier. In this case, it may be preferable to base inference on procedures that are more resistant and which specifically takes into account the fact that the assumed models used by the analysts are only approximate.

In general, the stability of parametric procedures generally requires strict adherence to the model assumptions, a condition which may be questionable in practice. Indeed, the assumed models are often chosen because they are consistent with the majority of the data, not because they fit exactly the data. Moreover, it is well-known that classical optimum procedures behave quite poorly under slight violations of the strict model assumptions. In order to produce statistical procedures that are

stable with respect to small changes in the data or to small model departures, robust statistical methods can be considered (see, among others, Huber (1981), Hampel *et al.* (1986), Maronna *et al.* (2006), Heritier *et al.* (2009)). Robust statistics is an extension of classical statistics that specifically takes into account the fact that the assumed parametric models used by the researchers are only approximate.

Typically, biostatisticians continue to use ad-hoc techniques to deal with outliers and underestimate the impact of model misspecifications. However, it is well-known that removing outliers with a simple data screening and then applying classical inferential procedures is not a simple and good way to proceed. First of all, in multivariate or highly structured data, it can be difficult to single out outliers, or it can be even impossible to identify influential observations. Second, in many cases it could be more efficient to down-weight instead of discarding certain observations. Rejecting outliers reduces the sample size, has effects on the distributional assumptions, and leads to underestimate dispersion. Finally, empirical evidence shows that good robust procedures behave quite better than techniques based on simple rejection of screened outliers.

There exists a great variety of approaches towards the robustness problem. Among these, procedures based on M-estimators (and gross error sensitivity) and high breakdown point estimators play an important and complementary role. The breakdown point of an estimator is the largest fraction of the data that can be moved arbitrarily without perturbing the estimator to the boundary of the parameter space: thus the higher the breakdown point, the more robust the estimator against extreme outliers. However, the breakdown point is not enough to assess the degree of robustness of an estimator. Instead, the gross error sensitivity gives an exact measure of the size of robusness, since it is a measure of the maximum effect an observation can have on an estimator. There are several books on robust statistics. Huber (1981), Hampel *et al.* (1986) and Maronna *et al.* (2006) are the main theoretical ones; see also Staudte and Sheather (1990). Rousseeuw and Leroy (1987) is more practical, and Heritier *et al.* (2009) focuses on techniques commonly used in biostatistics. Moreover, several libraries in the `R` software (R Development Core Team, 2009) have been implemented to perform robust analyses.

Despite the considerable developments from the theoretical point of view, in our experience at present robust methods are rarely used in medical and related fields. Indeed, biomedical scientific papers continue to use ad-hoc techniques to deal with outliers and model misspecifications. The aim of this paper is to fill the existing gap between theoretical robust techniques and the analysis of real datasets, and to encourage the dissemination of such inferential methods in the medical, genetic and, more generally, in the health sciences communities. To this end, in this paper we overview and illustrate the use of robust procedures for a wide variety of data structures which statisticians often encounter in practice, such as the well-known one-sample and two-sample t-tests, regression models, logistic regression, survival analysis and ROC curves. A related account may be found in Heritier *et al.* (2009), which gives many further examples. For reasons of space we have decided to focus on robust statistical tools which are more often used in medical research. Other important robust statistical methods, not considered here, involve robust principal components analysis (e.g. Rousseeuw and Leroy (1987), Maronna (2005), Hubert

et al. (2005)), clustering (e.g. Kaufman and Rousseeuw (1990), Cuesta-Albertos *et al.* (1997), Garcia-Escudero and Gordaliza (1999), Gallegos and Ritter (2005)), double clustering (Farcomeni, 2009), discriminant analysis (e.g. Hubert and Van Driessen (2004)), and mixed linear models (Copt and Victoria-Feser, 2006).

The outline of the paper is as follows. In Section 2 we set up the basics and describe robust procedures for univariate inference. In Section 3 we extend the approaches to scale and regression models. In Section 4 we describe an approach to robust logistic regression, and approaches to robust survival analysis are outlined in Section 5. Finally, in Section 6 we focus on the receiver operating characteristic curves (ROC), obtained from the response values of a diagnostic test based on a continuous diagnostic marker. Each section is concluded with a real data example illustrating the approaches. Finally, in Section 7 we briefly list `R` functions and packages implementing the reviewed procedures.

2 Background on basic robust procedures

It is well-known that the sample mean, and thus also the classical t-test or the ANOVA model, can be upset completely by a single outlier. In view of this, although the sample mean is the optimal estimator under the normal distribution, it can be substantially sub-optimal for distributions close to the normal. In clinical trials or in epidemiological studies, when the normality assumption is questionable, typically non-parametric methods are used and this may explain why they are popular for some types of analysis. However, power issues have to be taken into consideration. Indeed, robust methods aim to have high efficiency in a neighbourhood of the assumed statistical model. That is, it is often possible to calibrate robust procedures in order to achieve a pre-specified efficiency at the assumed model (e.g. 90-95%) and this means that the price to pay for the use of such tecniques with respect to parametric tests is a small loss in efficiency, on the contrary of rank-based methods.

The aim of this section is twofold. First, we introduce some basic concepts and measures of robust inference. Second, we illustrate robust versions of the one-sample and two-sample t-test, and, more generally, robust procedures for comparison of groups. We refer mainly to the general class of M-estimators for robust inference, and to the concept of influence function (see e.g. Hampel *et al.* (1986)).

2.1 Measuring robustness

Let us consider a sample $y = (y_1, \ldots, y_n)$ of size n, with independent and identically distributed components from a parametric model $F_\theta = F(y; \theta)$, with corresponding density $f(y; \theta)$, with $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$. Robust statistic aims at producing consistent and possibily efficient procedures, which are stable under small deviations from the assumed model. By slight model misspecification, we mean that the data-generating process lies in a neighborhood of F_θ, that is considered a sensible model for the problem under investigation. The notion of neighborhood can be formalized as $F_\varepsilon = (1 - \varepsilon)F_\theta + \varepsilon G$, where G is an arbitrary distribution and $0 \leq \varepsilon \leq 1$ (see Huber (1964)).

One way of assessing the robustness of a statistic $T = T(y)$, which can be represented (at least asymptotically) as a function of the empirical distribution function, is by means of the influence function (IF)

$$IF(x; T, F_\theta) = \lim_{\varepsilon \to 0} \frac{T(F_{\varepsilon,\Delta}) - T(F_\theta)}{\varepsilon} = \left. \frac{\partial T(F_{\varepsilon,\Delta})}{\partial \varepsilon} \right|_{\varepsilon=0} , \quad (1)$$

where $T(F_{\varepsilon,\Delta}) = T((1-\varepsilon)F_\theta + \varepsilon\Delta_x)$, with Δ_x point mass in x. The IF measures the local stability of a statistical procedure, since it describes the effect of an infinitesimal contamination at the point x. As a consequence, the linear approximation $\varepsilon IF(x; T, F_\theta)$ describes the asymptotic bias of T under single-point contaminations of the asssumed model distribution F_θ. The supremum of the IF, i.e. the gross-error sensitivity (GES), measures the worst influence on T and a desiderable robustness property is a finite GES, i.e. a bounded IF (B-robustness). Other robustness measures can be derived from the IF, such as the local-shift sensitivity, which measures robustness with respect to rounding effects (see Hampel *et al.* (1986)). Finally, note that the IF can also be used to evaluate the asymptotic covariance matrix of T, since $Var(T) = E_\theta(IF(x; T, F_\theta)IF(x; T, F_\theta)^\mathsf{T})$.

A second way to assess robustness is by considering the breakdown point (BP), which measures the robustness properties of a statistic T in a global sense. It is defined as the maximal amount of model misspecification an estimator can withstand before its bias becomes too large (infinite), i.e. it breaks down. In a formal definition, if the amount of model misspecification is related to the quantity ε in F_ε, BP can be defined as $\inf\{\varepsilon : \text{bias}(T, F_\theta, \varepsilon) = \infty\}$. If the GES of T is infinite, then its BP is nil. In practice, the BP is less used to assess robustness than the IF (and the related GES). Usually, the IF and the GES are first measured and then if they are bounded, the BP is evaluated in a second step. In view of this, a robust estimator with bounded IF can become useless in practice if its BP is too small.

Finally, a further measure for comparing different robust estimators is the rejection point (RP). It is mainly used in multivariate settings and it is defined as the distance from the center of the data such that points lying outside this distance have no influence on the asymptotic bias. Formally, for a symmetric F_θ centered at m, RP is defined as $\inf\{r > 0 : IF(x; T, F_\theta) = 0 \text{ when } \delta(x, m) > r\}$, where δ is a suitable distance measure. If an estimator has a finite RP, then points too far away from the center of the data receive a weight of zero.

2.2 Robust estimation and testing

A good compromise between robustness and efficiency can be obtained with the general class of M-estimators. Given a random sample y from F_θ, an M-estimator of θ is implicitly defined as a solution $\widehat{\theta}$ of the unbiased estimating equation (see e.g. Hampel *et al.* (1986))

$$\Psi_\theta = \sum_{i=1}^{n} \psi\left(y_i; \theta\right) = 0 , \quad (2)$$

where $\psi(\cdot)$ is a suitable function. If $\psi(\cdot)$ is the score function, then the maximum likelihood estimator (MLE) is obtained. Moreover, also well-known simple robust

estimators of location and scale, such as the median, the median absolute deviation (MAD) and the trimmed mean, belong to the class of M-estimators.

Under broad regularity conditions, an M-estimator, as the MLE within a parametric model, is consistent and approximately normal with mean θ and variance

$$V(\theta) = B(\theta)^{-1}\Omega(\theta)(B(\theta)^{-1})^{\intercal} \ , \tag{3}$$

where $B(\theta) = -E_\theta(\partial\Psi_\theta/\partial\theta^{\intercal})$ and $\Omega(\theta) = E_\theta(\Psi_\theta\Psi_\theta^{\intercal})$. Thus, confidence intervals (or testing hypotheses) can be performed in a usual way by using a consistent estimate of the asymptotic variance $V(\theta)$. Robust inference for θ is typically based on the Wald-type statistic

$$W_\psi^2 = (\widehat{\theta} - \theta)^{\intercal} V(\widehat{\theta})^{-1}(\widehat{\theta} - \theta) \ , \tag{4}$$

which has an asymptotic χ_p^2 distribution. For instance, in the special case of $p = 1$, the Wald-type statistic

$$t(\theta) = \frac{(\widehat{\theta} - \theta)}{V(\widehat{\theta})^{1/2}} \tag{5}$$

may be used to set confidence intervals or compute a p-value for θ. Suppose, for instance, that it is desired to use the statistic $t(\theta)$ to test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta > \theta_0$. As a large positive value of $t(\theta_0)$ relative to the $N(0,1)$ distribution will give evidence against H_0, the corresponding p-value is $1-\Phi(t(\theta_0))$, where $\Phi(\cdot)$ denotes the standard normal distribution function. Likewise a confidence interval may be constructed using those values of θ most consistent with this distribution. The resulting interval is $(\widehat{\theta} \pm z_{1-\alpha/2}V(\widehat{\theta})^{1/2})$, with z_α denoting the α-quantile of the $N(0,1)$ distribution, which can be easily computed. In Section 3.3 also likelihood ratio-type tests derived from M-estimators will be discussed.

It can be shown that the IF of an M-estimator is given by

$$IF(x; \widehat{\theta}, F_\theta) = B(\theta)^{-1}\psi(x;\theta) \ . \tag{6}$$

Since the IF of an M-estimator is proportional to its ψ-function, an M-estimator $\widehat{\theta}$ is B-robust if and only if $\psi(x;\theta)$ is bounded. This is a powerful result since it suffices to choose a bounded ψ-function to obtain a robust M-estimator. In general, MLEs have unbounded IF. Finally, notice that the stability of the test level and of the power under small departures from the assumed model is proportional to the IF of the test statistic (see e.g. Heritier and Ronchetti (1994)). As a consequence, a test statistic of the form (4) with bounded IF guarantees the local stability of the level and of the power with respect to small deviations from the assumed model.

Robust versions of the classical t-test and test statistics for comparing means among groups can be simply obtained for suitable choices of the ψ-function. In particular, robust t-tests are usually derived under the assumption of the Huber's estimating functions for location and scale, and then using the Wald-type statistics (4) and (5). The well-known Huber's ψ-function for location is given by $\psi(y;\theta) = \psi_H(y - \theta; k)$, with

$$\psi_H(x;k) = \max(-k, \min(k, x)) \ , \tag{7}$$

where k is a constant the user specifies in order to control the degree of robustness of the procedure, taking into account the loss of efficiency he is prepared to accept at the model in exchange to robustness. Its limit as $k \to 0$ is the median, and as $k \to \infty$ is the mean. The constant k is usually chosen using efficiency arguments, i.e. so that the ratio between the variance of the MLE at F_θ and of the robust estimator achieves a given value, typically 90-95% (see also Section 2). The value $k = 1.345$ gives 95% efficiency at the normal model. For the scale parameter the Huber's Proposal 2 estimate can be considered, with ψ-function given by $\psi_H(y;k)^2 - k_1$, with k_1 suitable constant which garantees consistency at the assumed model. The IF of the Huber's estimate for location can be seen as a compromise between the IF of the mean and of the median: it remains identical to the mean's IF in the middle and then is truncated symmetrically beyond a certain threshold. Since the form of the IF has implications on the efficiency of the corresponding estimator (see Section 2.1), the Huber's estimator is more efficient than the median.

Alternatively, it is possible to resort to the ψ-function of the optimal B-robust estimator (OBRE), i.e. the M-estimator with maximal efficiency amog all M-estimators with bounded IF, measured in an appropriate metric (see Hampel *et al.* (1986)). Or, in order to have M-estimators with larger BP than the Huber's or the OBRE estimators, the so-called redescending ψ-functions can be considered, that are ψ-functions that become nil for large values of their arguments or such that their RP is finite. The most popular redescending estimator is the Tukey's bisquare proposal (see Huber (1981), Hampel *et al.* (1986) and Venables and Ripley (2002)).

2.3 An example: Rehabilitation therapies on arm motor performance

The dataset considered here contains measurements on the functional independence measure (FIM) scores on poststroke subjects (see Piron *et al.* (2010)). FIM scores have been measured pre and after a physical therapy treatment received in the early period after stroke and on two groups of subjects: 27 patients treated with a rehabilitation technique in a virtual environment (cases) and 20 patients treated with a conventional theraphy (controls). Figure 1 illustrates the two boxplots of the FIM differences scores (differences between post-treatment and pre-treatment measurements) in the two groups of patients. The boxplot is a useful plot since it allows to identify possible outliers and to look at the overall shape of a set of observed data. While for the second group of patients (controls) the normal distribution can be assumed, this assumption may be questionable for the first group (cases).

Table 1 gives the estimated means (and standard errors) in the two groups, and the Huber's estimators for location (and the MADs), assuming $k = 1.345$ under the normal model. Note that while the sample mean in the first group is larger than the sample mean in the second group, when robust estimators are considered, the Huber's estimator is larger in the second group (as well as the median). Indeed, the mean is not robust and its value depends strongly on the extreme observations in case subjects. Moreover, a relevant effect of outlying observations is on the classical estimate of the scale parameter. Table 1 also gives the value of the two-sample t-test, with its p-value, and the confidence interval for the difference of the two means of the two groups. Also the value of the robust test, and its p-value, and the robust

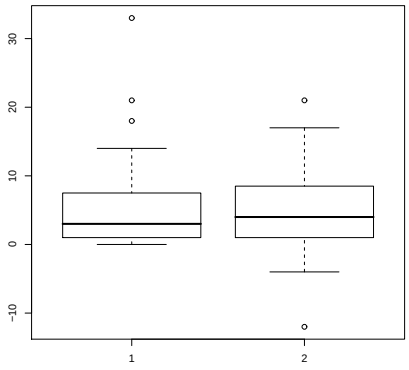

Figure 1: Boxplot of the differences pre-post FIM scores in case (1) and control patients (2).

confidence interval for the difference of the two means are given. Note that, although both the test statistics do not reject the null hypothesis of 'no difference between the mean FIM across the two groups', the robust test, as well as the robust CI, is quite different from the classical t-test.

	$Estimate$ $location(scale)$	$test$ (IC)	$p-value$
Mean cases	5.63 (7.74)	0.42	0.67
Mean controls	4.70 (7.18)	(-3.53,5.39)	
Huber cases	3.90 (4.44)	-1.53	0.12
Huber controls	4.49 (5.18)	(-4.92,0.60)	

Table 1: Summary of two-groups comparison for FIM data.

The lack of stability of many classical inferential procedures also affects the corresponding p-values. As a further illustration of this, let us focus on the FIM differences between post-treatment and pre-treatment measurements in control patients, and suppose it is of interest to test the null hypothesis H_0 of 'no effect of the treatment'. Under the assumption of normality, the null hypothesis H_0 is typically tested by means of a one-sample Student t-test, and let us consider also the Wilcoxon rank-sum test and the robust test based on the Huber's estimator. For the observed data, all these test statistics return p-values smaller than 0.01. To illustrate the instability of the p-value associated to the t-test, a sensitivity analysis has been considered: it consists in replacing one observation of the sample by an arbitrary value which moves from -20 to 20 and thus in plotting the values of the p-values. The plot of the t-test, Wilcoxon test and Huber's test p-values is given in Figure 2. The variation in the t-test's p-value shows its instability. In contrast, the Huber's test is a safer procedure that returns a stable p-value around 0%-1%. The Wilcoxon rank-sum test has an intermediate behaviour with a stable p-value when

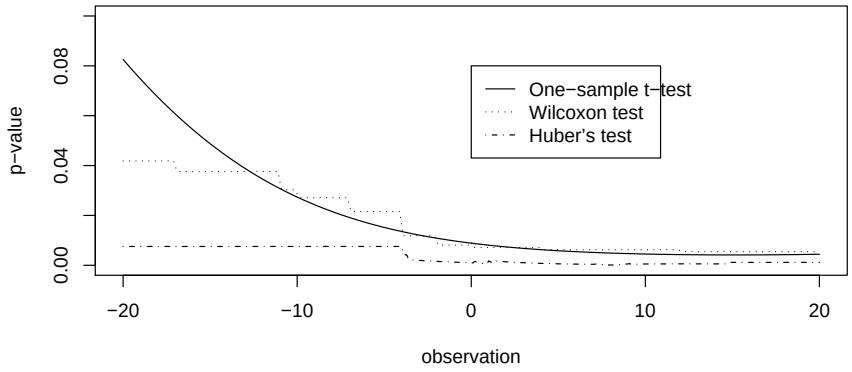

Figure 2: Sensitivity analysis of the p-values for the t-test, Wilcoxon test and Huber's test when an observation in differences pres-post FIM scores in control patients is changed.

the observation is greater than -2.

3 Robust inference in scale and regression models

Scale and regression models are probably the most widely used statistical tools in medical and related sciences, since they represent the simplest models to describe relationships between a continuous response variable and a set of explanatory variables. They include as special cases the well-known t-statistic and the ANOVA models for comparisons across categories of subjects. The aim of this section is to give an overview of classical and more recent robust procedures to fit, test and check a scale and regression model. The main references for the arguments discussed here are, among others, Huber (1981), Li (1985), Hampel *et al.* (1986), Venables and Ripley (2002), Maronna *et al.* (2006), Heritier *et al.* (2009), and references therein.

A scale and regression model is given by

$$y_i = x_i^{\mathsf{T}} \beta + \sigma \varepsilon_i \ , \quad i = 1, \ldots, n \ , \tag{8}$$

where $y = (y_1, \ldots, y_n)$ is the response, x_i is a known p-variate vector of regressors, $\beta \in \mathbb{R}^p$ $(p > 1)$ is an unknown p-vector of regression coefficients, $\sigma > 0$ a scale parameter and ε_i are independent and identically distributed random variables according to a known density function $p_0(\cdot)$. In many applications, it is assumed that $p_0(x) = \phi(x)$, where $\phi(x)$ is the standard normal density. In this situation, exact confidence intervals and tests based on the t and F distributions may then be computed using classical least squares estimates (LSE) or MLEs, available in any statistical packages. More generally, models of form (8) are characterized by a linear predictor and possibly non-normal errors. Applications may be found in many areas of health sciences, such as epidemiology and anthropology. Popular choices

for $p_0(\cdot)$ include the Student's t, extreme value, logistic, skew-normal and Cauchy distributions (see, e.g., Brazzale *et al.* (2007)).

Classical estimates for the parameters of model (8) are optimal, in the sense that they are the most efficient (and consistent), but they are extremely sensitive to outliers. Indeed, it is well-known that a few atypical observations or small deviations from the assumed $p_0(\cdot)$ can have a large influence on both the LSE and MLE, leading to biased estimators and hence wrong interpretations of the fitted models. For instance, if $p_0(x)$ is a Student distribution with three degrees of freedom, the MLE under the correct assumption on $p_0(\cdot)$ is better than the LSE under the normal model: the efficiency loss of the LSE with respect the MLE can be as large as 50%. Biases introduced by even small model deviations may also have consequences on the evaluation of the residuals, and hence on the check of the model. It could be tempting to remove outliers to avoid possible bias with classical estimators, using for instance graphical analyses, residual analyses or more sophisticated diagnostic tools, such as the Cook and Weisberg (2007) distance or forward searches for outliers detection (Atkinson and Riani, 2000). However, although apparently simple, these strategies can be not only impractical, but also misleading; see for instance Welsh and Ronchetti (1985).

There are a number of ways to perform robust inference in regression models. In regression there are two possible sources of errors: the observations in the response and/or the p-variate vector of regressors (see Figure 3 for simple regression). Some robust methods in regression only consider the first source of outliers (outliers in y-direction), and in some situations of practical interest errors in the regressors can be ignored (such as in most bioassay experiments). It is well-known that outliers in y-direction have infuence on classical estimates of β, but in particular on the classical estimates of the error variance. The presence of outliers in y-direction emerges typically in the graphical analysis of the residuals of the model.

Outliers in x-direction can also be outlying in the y-direction, consequently having a small influence. Leverage points are outliers in the x-direction which are not in the y-direction. These kind of points can be very dangerous since they are typically very influential. In low dimensions, graphical analyses can be used for their detection, but this approach can be very difficult with high-dimensional data, since cases with high leverage may not stand out in least squares residual plots.

In the scale and regression framework, different classes of estimators and related inferential procedures can be identified. The most commonly used robust methods to deal with outliers in y-direction are Huber-type estimators. Where there is a moderate percentage of multivariate outliers in the x-direction, or leverage points, bounded influence estimators may be preferable. When the proportion of outliers is very high (up to 50%), we can use the high breakdown point estimators (Sections 3.1 and 3.2).

When interest is in testing the significance of the regression parameters, Wald-type procedures based on robust estimation can be used (see e.g. Maronna *et al.* (2006)), but it may be preferable to resort to robust likelihood functions derived from the estimating equations of robust estimators (Section 3.3). Finally, to obtain a measure for the goodness of fit of the model a robust version of the coefficient of determination may be used (Section 3.4).

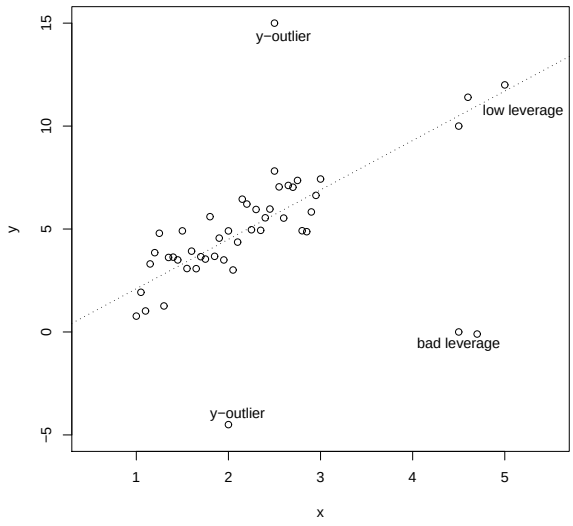

Figure 3: Different types of outliers in simple regression models.

In this section, some theoretical robust methods in linear regression are inevitably skipped. For instance, situations with heteroscedastic models (Bianco *et al.* (2000), Carroll and Ruppert (1982), Giltinan *et al.* (1986)), or models for repeated measures (Welsh and Richardson (1997), Copt and Victoria-Feser (2006), Copt and Heritier (2007)) are not considered.

3.1 M-estimation

The development of robust regression started with M-estimators (Huber (1973), Huber (1981)). A wide class of robust M-estimators for the regression coefficients, which have the advantage of combining robustness with efficiency under the regression model with normal errors, is defined by estimating functions of the form

$$\Psi_\beta = \sum_{i=1}^{n} s(x_i)\psi_\beta(r_i) = \sum_{i=1}^{n} w_\beta(r_i) r_i x_i \ , \tag{9}$$

where $r_i = (y_i - x_i^\mathsf{T}\beta)/\sigma$ denotes the i-th residual, $s(\cdot)$ and $\psi_\beta(\cdot)$ are given functions, and $w_\beta(\cdot)$ are suitable weights related to $\psi_\beta(\cdot)$, which make $w_\beta(r_i)r_i$ bounded (see Hampel *et al.* (1986), Ch. 6). In particular, when $s(x) = 1$ and $\psi_\beta(\cdot) = \psi_H(\cdot; k)$ we obtain Huber's estimator for regression; in this case, $w_\beta(r) = w_\beta^H(r) = \psi_H(r; k)/r = \min(1, k/|r|)$. In applications, the weights $w_\beta(r_i)$, $i = 1, \ldots, n$, are worth looking at because they automatically define the observations that have been considered by the estimator as more or less far from the bulk of data. In particular, one can determine approximately the amount of contamination. To estimate the scale parameter σ of model (8), an M-estimating equation can be taken, i.e. $\Psi_\sigma = \sum_{i=1}^{n} \psi_\sigma(r_i) = 0$, where $\psi_\sigma(\cdot)$ is an appropriate function. A popular choice is the MAD of the residuals.

Huber-type estimates are not robust with respect to bad leverage points, i.e. outliers in x-direction, since the Huber's weight $w_\beta^H(r)$ only controls for extreme residuals. To obtain estimates which are robust against any type of outliers, the bounded-influence Mallow's class can be considered, defined by a suitable weight function $0 \leq s(x) \leq 1$ and $\psi_\beta(\cdot) = \psi_H(\cdot; k)$ in (9) (see, for instance, Hampel *et al.* (1986), Maronna *et al.* (2006)). The choice $s(x) = 1/||x||$ defines the Hampel-Krasker M-estimator for regression. With continuous regressors, a classical approach is to compute a robust Mahalanobis distance, given by

$$d_i = \sqrt{(x_i - \widehat{\mu})^\top \widehat{\Sigma}^{-1}(x_i - \widehat{\mu})} \ . \tag{10}$$

In (10), $(\widehat{\mu}, \widehat{\Sigma})$ are robust estimates of multivariate location and scatter, such as the minimum covariance determinant estimator (MCD) of Rousseeuw (1984); see also Hubert *et al.* (2008). The weights can thus be defined as $s(x_i) = 1$ if $d_i^2 \leq \chi^2_{p,0.975}$, and $s(x_i) = 0$ otherwise. A further possibility to choose a weighting scheme in (9), which can be used also for categorical regressors, is based on the leverages h_{ii}, i.e. on the diagonal elements of $H = X(X^\top X)^{-\top} X^\top$, which are well-known in regression diagnostic (see e.g. Cook and Weisberg (2007)). The weights based on the leverages are thus $s(x_i) = \sqrt{1 - h_{ii}}$, $i = 1, \ldots, n$, and compensate for the extreme responses not captured by the residuals.

The Hampel-Krasker estimator is optimal in the sense that it is the M-estimator which minimizes the trace of the asymptotic covariance matrix under the constraint that it has a bounded IF. This estimator is however optimal for $p_0(x) = \phi(x)$. In the context of non-normal linear regression models, Carroll and Ruppert (1988) discuss general procedures to obtain the OBRE. For a given bound k on the IF, the OBRE for $\theta = (\beta, \sigma)$ is implicitly defined by

$$\sum_{i=1}^{n} \psi(y_i; \theta) = \sum_{i=1}^{n} \{\ell_\theta(\theta; y_i) - a(\theta)\} w_i = 0 \ , \tag{11}$$

where $\ell_\theta(y; \theta) = \partial \log p(y; \theta)/\partial\theta$, $w_i = \min(1, k/||\ell_\theta(\theta; y) - a(\theta)||_{A(\theta)})$, $||x||_A = [x^\top A^{-1} x]^{1/2}$, and the $p \times p$ matrix $A(\theta)$ and the $p \times 1$ vector $a(\theta)$ are defined by the conditions $A(\theta) = E(\psi(y; \theta)\psi(y; \theta)^\top)/n$ and $E(\psi(y; \theta)) = 0$. The interpretation of this result is quite simple. For efficiency reasons, the OBRE has to be as similar as possible to the MLE for the values of y in the bulk of data, i.e. at noninfluential values of y: therefore, its ψ-function equals the score function ℓ_θ for those values. On the other hand, since the IF is proportional to the ψ-function, in order to obtain a bounded IF one has to truncate ℓ_θ where the bound k is exceeded. This is achieved by means of the weights w_i, $i = 1, \ldots, n$, that automatically define the observations that have been considered by the OBRE as more or less far from the bulk of data. The matrix $A(\theta)$ and the vector $a(\theta)$ can be viewed as Lagrange multipliers for the constraints resulting from a bounded IF and Fisher consistency. The constant k is the bound on the IF and can be interpreted as the regulator between robustness and efficiency: for a lower k one gains robustness but loses efficiency, and viceversa for a higher k. Typically, the choice of k depends in general on the model. To compute the OBRE an iterative algorithm is required (see Bellio (2007)).

3.2 Other robust estimators

The IF is not the only important robustness measure. Another key concept is the BP (see Section 2.1). We remember that the BP of the Huber's estimator and LSE is 0%, and in general it cannot exceed $1/p$ for other robust M-estimators (that is, it decreases with increasing dimension where there are more opportunities for outliers to occur). The more resistant M-estimators with respect to leverage points are the ones with redescending ψ-functions, i.e. ψ-functions that can become nil when the residuals are too large. One example is the Tukey's biweight function (see, e.g., Hampel *et al.* (1986)). However, redescending estimating functions may have multiple roots, and this may considerably complicate the computation of the estimates. In such cases it is important to choose a good starting point and iterate carefully.

Several alternative high breakdown point and computationally efficient estimators of regression have been proposed. The first high-breakdown regression estimator to become popular was the least median of squares (LMS), defined as $\widehat{\beta}_{lms} = \min med(y_i - x_i^{\intercal}\beta)^2$. This fit is very resistant and needs no scale estimate. The BP of the LMS estimator is 50%, but this estimator is highly inefficient when the central model is the normal one. Another proposal is the least trimmed squares (LTS) estimate, defined as $\widehat{\beta}_{lts} = \min \sum_{i=1}^{k} r^2_{[i]}$, where $r^2_{[1]} \leq r^2_{[2]} \leq \ldots \leq r^2_{[n]}$ are the ordered squared residuals, and k is the largest integer such that $k \leq n/2 + 1$. This is equivalent to find the k-subset with smallest least squares objective function, which resembles the definition of the MCD. This estimator is more efficient than LMS, having BP 50%. When using LTS regression, the scale parameter σ can be estimated as $\widehat{\sigma}_{lts} = c_k\sqrt{r^2_{[i]}/k}$, where c_k makes $\widehat{\sigma}_{lts}$ consistent and unbiased at $p_0(x) = \phi(x)$. Also $\widehat{\sigma}_{lts}$ is highly robust and therefore regression outliers may be identified by the standardized LTS residuals $r_i/\widehat{\sigma}_{lts}$. Finally, we remember that high breakdown procedures do not usually provide standard errors. However, these can be obtained by a data-based simulation, such as a bootstrap.

It is possible to combine the resistance of high breakdown estimators with the efficiency of M-estimation. The resulting estimators are called MM-estimators (see Yohai (1987), and Marazzi (1993)) and have an asymptotic efficiency as close to one as desired, and simultaneously breakdown point 50%. Formally, the MM-estimate consists in solving an M-type redescending estimating equation, using a consistent estimator with high breakdown (even with low efficiency) as starting point. Instead of the redescending estimating equation, other ψ-functions may be choosen. An alternative approach to derive robust estimates against any type of outliers, is to fit linear models using the weighted likelihood (Markatou *et al.* (1998), Agostinelli and Markatou (1998)): the resulting estimator is robust against the presence of bad leverage points too. The weighted likelihood methodology consists in contructing a weight function $w(\cdot)$ that depends on the data y and on the distribution of the assumed parametric model. The estimators of the parameters are then obtained as solutions of a set of estimating functions of the form

$$\sum_{i=1}^{n} w(y_i)\ell_\beta(\beta; y_i) = 0 \ ,$$

where $\ell_\beta(\beta; y_i)$ is the score function of observation y_i. There are several proposals for the weight function $w(\cdot)$, depending on a choosen distance. Note that the robust estimates can be interpreted as redescending estimates with adaptive ψ-function, that is the shape of the ψ-function depends on the observed data.

3.3 Testing

Large sample Wald-type tests and confidence regions for β and σ can be constructed in a standard way using an estimate of the asymptotic covariance matrix of the estimators (see Section 2.2). Let $SE(\widehat{\beta}_j) = (\widehat{\text{var}}(\widehat{\beta})_{jj})^{1/2}$, with $\widehat{\text{var}}(\widehat{\beta})$ suitable estimate for the asymptotic covariance matrix. Then, the Wald-type test statistic for testing significance of regression parameters is simply given by the ratio

$$t_j = \frac{\widehat{\beta}_j}{SE(\widehat{\beta}_j)} \ .$$

In view of the asymptotic normality of M-estimators, the p-value for the Wald-type robust test is obtained by comparing t_j with the standard normal distribution.

For classical likelihood-based procedures it is well-known that the use of likelihood ratio statistics leads to inferential results that are more accurate than those pertaining to Wald-type statistics, especially when the sample size is small. A possible alternative to Wald-type procedures is to derive suitable robust likelihoods from the estimating equations that define robust estimators. Their robustness properties are driven from the corresponding M-estimators (Heritier and Ronchetti, 1994). Then, these robust likelihoods can be used to define robust likelihood ratio-type tests.

Starting from an unbiased estimating function for β, a quasi-likelihood for β is given by McCullagh (1991)

$$\ell_Q(\beta) = \int_c^\beta A(t)\, \Psi_\beta(y; t) dt, \tag{12}$$

with c arbitrary constant, and $A(\beta)$ such that $A(\beta)^\intercal = \Omega(\beta)^{-1} B(\beta)$. For setting quasi-likelihood confidence regions or for testing hypotheses, the quasi-likelihood ratio statistic $W_Q(\beta)$, with the standard χ^2 distribution, may be used. When β is scalar, $\ell_Q(\beta)$ usually exists and is easy to derive, but, for $p > 1$, (12) does not exist in general. A sufficient condition is that the matrix $\partial \Psi_\beta / \partial \beta^\top$ has to be symmetric. The problem of nonexistence of (12) may be overcome when interest is on a scalar component of β. In particular, suppose that the parameter of interest is β_j and let λ be the remaining parameters of the model. In this setting, it is possible to define a quasi-profile likelihood for β_j (Adimari and Ventura (2002b), Bellio *et al.* (2008)), given by

$$\ell_{QP}(\beta_j) = \int_c^{\beta_j} w(t, \tilde{\lambda}_t) \Psi_\tau(y; t, \tilde{\lambda}_t) dt, \tag{13}$$

where $w(\cdot)$ is a suitable correction term. The corresponding quasi-profile likelihood ratio statistic $W_{QP}(\beta_j)$ is approximately χ^2_1 distributed.

If in (12), or in (13), the adjustment term $A(\beta)$ is not considered, then the corresponding quasi-likelihood ratio test statistic converges in distribution to $\sum_{j=1}^{d} \delta_j \chi^2_1$, where the δ_j are the eigenvalues of the matrix $\Omega(\beta)B^{-1}(\beta)$ (see, also, Heritier and Ronchetti (1994), Markatou and Hettmansperger (1990b), Markatou and Hettmansperger (1990a), Hanfelt and Liang (1995)). To achieve the usual asymptotic distribution, the quasi-likelihood ratio test statistic may be adjusted with the scale correction $\mathrm{tr}(\Omega(\tilde{\beta})B^{-1}(\tilde{\beta}))/p$. When interest focuses on a scalar β_j, the corresponding quasi-profile likelihood ratio statistic converges in distribution to a χ^2_1 as $n \to \infty$, and, as pointed out in Bellio *et al.* (2008)), this statistic is actually asymptotically equivalent to $W_{QP}(\beta_j)$.

The last approach considers the empirical likelihood ratio statistic (Owen, 2001) for β, given by $W_E(\beta) = -2 \log R_E(\beta) = \max_{p_i} \prod_i np_i$, where the p_i-weights satisfy $p_i > 0$, $\sum_i p_i = 1$ and $\sum_i \psi(y_i, \beta)p_i = 0$. Its main appeal is that only unbiasedness of the estimating function is required to obatain a standard asymptotic χ^2_p distribution for the empirical likelihood ratio statistic $W_E(\beta)$. When inference focus on β_j, a profile version of $W_E(\beta)$ can be computed, given by $W_{EP}(\beta_j) = -2 \log \{\sup_\lambda R_E(\beta_j, \lambda)\}$, which is asymptotically distributed as a χ^2_1.

For applications of quasi- and empirical likelihoods for robust inference in linear models see, among others, Markatou and Hettmansperger (1990b), Markatou and Hettmansperger (1990a), Adimari and Ventura (2002b), Adimari and Ventura (2002a), Bellio *et al.* (2008), and Heritier *et al.* (2009) (Sect. 3.3.3).

3.4 Model checking

As in standard regression analyses, also when robust estimation is used, it may be important to check the model assumptions, evaluating for instance the approximate normality of the estimated standardized residuals by a normal qq-plot. Another diagnostic plot of a robust fit of a regression model can be considered, given by the plot of the fit weights of the robust estimator. See McKean *et al.* (1993) for the use and interpretability of the residual plots for a robust fit.

The fit weights of the robust estimator can also be used to define a robust version of the well-known coefficient of determination R^2, which is a measure for the goodness of fit of the model in the classical linear regression model. Indeed, when robust inference on the regression model is performed, it is possible to use a measure of goodness of fit by means of a robust version of R^2 (see Heritier *et al.* (2009)), given by

$$R^2_r = \left(\frac{\sum_{i=1}^{n} w_i (y_i - \bar{y}_w)(\widehat{y}_i - \bar{\widehat{y}}_w)}{\sqrt{\sum_{i=1}^{n} w_i (y_i - \bar{y}_w)^2 \sum_{i=1}^{n} w_i (\widehat{y}_i - \bar{\widehat{y}}_w)^2}} \right)^2 , \tag{14}$$

where $\bar{y}_w = (1/\sum w_i) \sum w_i y_i$, $\bar{\widehat{y}}_w = (1/\sum w_i) \sum w_i \widehat{y}_i$ and the weights w_i are the ones produced by the robust regression estimator used in the statistical analysis.

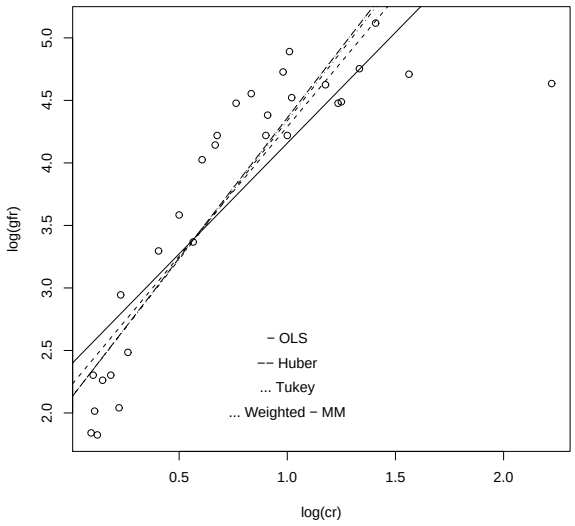

Figure 4: `GFR data`: LSE fit and several robust fits.

3.5 Example: Glomerular Filtration Rate (GFR) data

This dataset contains measurements of the glomerular filtration rate (`gfr`) and serum creatinine (`cr`); it has been discussed also in Heritier *et al.* (2009). The `gfr` is the volume of fluid filtered from the renal glomerular capillaries into the Bowman's capsule per unit of time and, clinically, it is used to determine renal function. To estimate `gfr` an endogenous molecule, synthsized in the body, may be used, that is the serum creatinine (`cr`).

Let us start the analysis with the simple linear relationship between $y = \log(\texttt{gfr})$ and $x = \log(\texttt{cr})$. The data are plotted in Figure 4 together with the LSE fit and several fitted robust regression lines (Huber's, Tukey's biweight, weighted likelihood and MM-estimation). From Figure 4 we can note that there are some differences between robust estimates and the LSE. Note that the LSE and Huber's estimated models are the more sensible to the two observations which looks extremes with respect to the linear regression model (outliers in x-direction). On the contrary, the more high breakdown point estimators behave quite similarly.

Figure 5 gives four possible diagnostic plots based on the classical LSE fit: the plot of the residuals versus the fitted values, the normal qq-plot of the residuals, the Cook's distance plot and the Cook's distance statistic versus $h_{ii}/(1 - h_{ii})$. These plots give different information about the observations. In particular, there is at least a leverage observations (which is also associated to a large residual). Figure 6 gives four possible diagnostic plots based on the robust Tukey's fit. Note that there are some observations with lower fitted weights which were not identified by the classical Cook's statistics, and that the extreme observation is rejected.

It may be useful to compare the different weights in the robust estimations

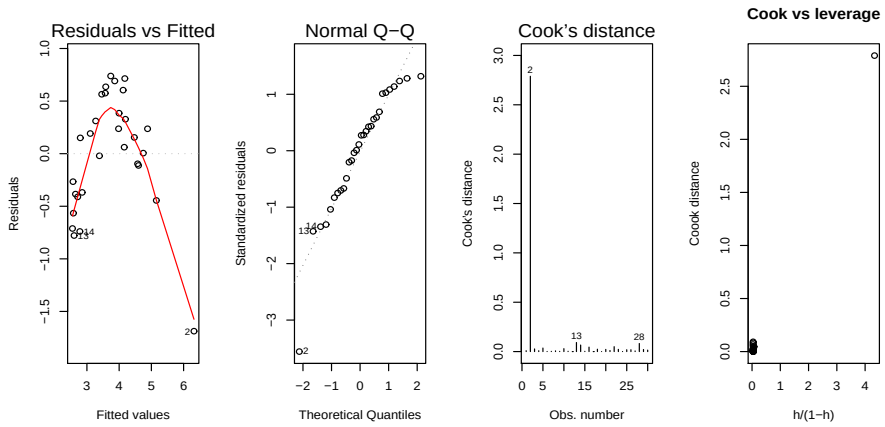

Figure 5: `GFR data`: diagnostic plots for LSE estimates.

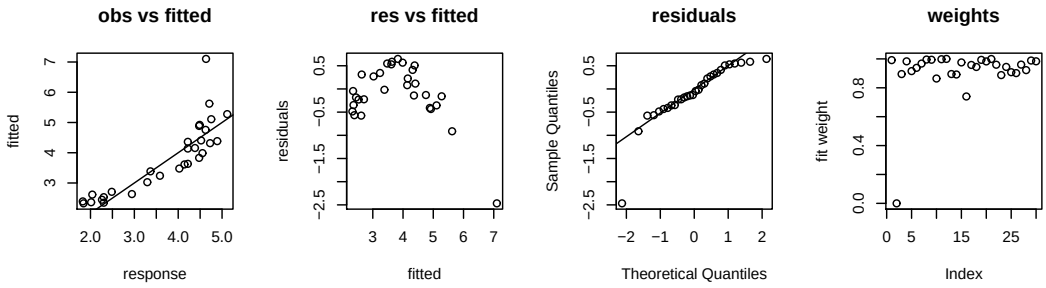

Figure 6: `GFR data`: diagnostic plots for robust estimates.

visualizing them (see Figure 7). In this way, the several ways of treating the data can be observed: bounding its influence and smooth rejection. Moreover, these weights can be used to determine approximately the amount of contamination and to compute the robust version of R^2.

Let us consider a more complex model for these data. In particular, as in Rule *et al.* (2004), let us consider the model

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 z + \varepsilon \ ,$$

where z is the age on the patients. Table 2 gives the different estimated values, together with the correspong p-values for significance testing based on Wald statistics. The conclusions of the fits are quite similar since the variable `age` is not significant in either model, while x and x^2 are always significant. Moreover, LSE and Huber's fit are very similar. On the contrary, LSE and weighted likelihood (or MM) estimates

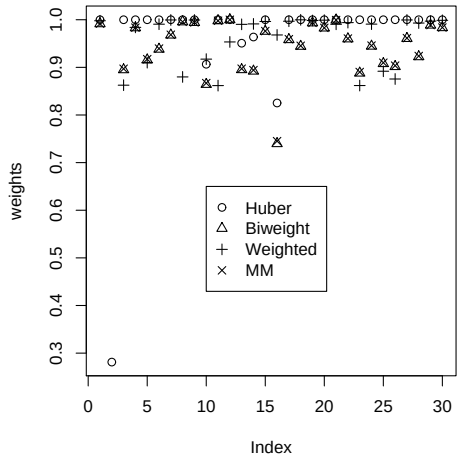

Figure 7: `GFR data`: weights of different robust estimates.

	LSE		Huber		Weighted		MM	
	estimate (se)	p-value	estimate (se)	p-value	estimate (se)	p-value	estimate (se)	p-value
β_1	1.80 (0.25)	$< 10^{-4}$	1.83 (0.27)	$< 10^{-4}$	1.56 (0.25)	$< 10^{-4}$	1.56 (0.25)	$< 10^{-4}$
β_2	4.27 (0.27)	$< 10^{-4}$	4.25 (0.29)	$< 10^{-4}$	5.02 (0.41)	$< 10^{-4}$	5.04 (0.27)	$< 10^{-4}$
β_3	-1.38 (0.14)	$< 10^{-4}$	-1.37 (0.15)	$< 10^{-4}$	-1.92 (0.27)	$< 10^{-4}$	-1.94 (0.14)	$< 10^{-4}$
β_4	-0.003 (0.003)	0.43	-0.003 (0.003)	0.429	-0.001 (0.003)	0.70	-0.001 (0.004)	0.73
$\widehat{\sigma}$	0.27		0.33		0.25		0.29	
R_r^2	0.943		0.943		0.952		0.88	

Table 2: `GFR data`: Estimates (and se) regression parameters.

give different results on both the estimated values of the regression coefficients and in terms of R_r^2. Figure 8 gives a graphical inspection useful to identify those observations with weights less than one. Note that the different robust procedures tend to downweight the observations in a different manner.

Diagnostic plots for the weighted likelihood fitted model have been considered (see Figure 9), in order to evaluate this fit. Similar plots can be obtained also for the other robust estimators. We note that the lower weight in the weighted likelihood fit corresponds to the observation with maximum value of the Cook's distance statistic.

4 Robust logistic regression

Departures from model assumptions and malicious observations can lead to problems also in estimation in generalized linear models (GLM). Here, for reasons of space, we focus only on binary logistic regression, where y_i can only take two values: zero and one. The model can be expressed as

$$\log\left(\frac{\Pr(Y_i = 1|x_i)}{1 - \Pr(Y_i = 1|x_i)}\right) = x_i^{\mathsf{T}}\beta, \tag{15}$$

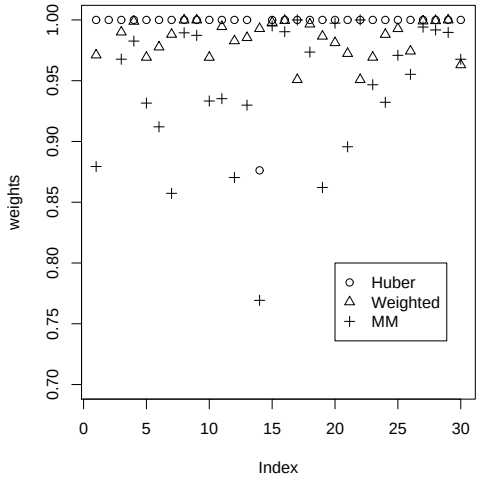

Figure 8: `GFR data`: weights of different robust estimates in the complex model.

where we have used the popular *logit* link. See, for instance, McCullagh and Nelder (1989) for details on the logistic regression model.

The classical MLE parameter estimates may break down due to leverage points, or to misclassification in the response (a zero instead of a one, or vice versa). This second case corresponds to a surprising scenario in which the covariates recorded for a misclassified subject, albeit not outlying in the x-direction, would clearly indicate the opposite outcome (i.e., the estimated probability of a zero is very low, but a zero is observed; or vice versa).

Many approaches to robust estimation for the logistic regression model have been proposed; see, e.g., Pregibon (1982), Copas (1988), Morgenthaler (1992), Carroll and Pederson (1993), Bianco and Yohai (1997), Markatou *et al.* (1997), Victoria-Feser (2002). Note that there are also methods derived for robust estimation in the entire GLM class, as for instance the OBRE by Künsch *et al.* (1989).

Here we review a popular approach due to Cantoni and Ronchetti (2001), who develop a Mallows-type estimator based on a modification of the system of estimating equations derived from the quasi-likelihood estimator (Wedderburn (1974), Heyde (1997)). This latter estimator is the solution of the system of estimating equations

$$\sum_{i=1}^{n} x_i^{\mathsf{T}} (y_i - \mu_i) \sqrt{V_i} = 0 \ ,$$

where $\mu_i = e^{x_i^{\mathsf{T}}\beta}/(1 + e^{x_i^{\mathsf{T}}\beta})$, $V_i = \mu_i(1 - \mu_i)$, $i = 1, \ldots, n$.

The Cantoni and Ronchetti (2001) approach uses an appropriate weighting scheme and the Huber's $\psi_H(\cdot; k)$ function as follows

$$\sum_{i=1}^{n} w(x_i) x_i^{\mathsf{T}} (\psi_H(r_i; k) - a(\mu_i)) \sqrt{V_i} = 0 \ , \tag{16}$$

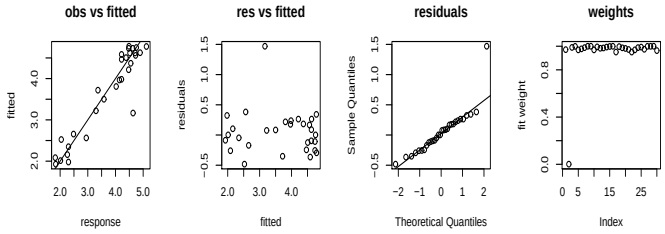

Figure 9: `GFR data`: diagnostic plots for weighted likelihood estimates.

where

$$a(\mu_i) = \psi_H\left((1-\mu_i)/\sqrt{V_i}; k\right)\mu_i + \psi_H\left(-\mu_i/\sqrt{V_i}; k\right)(1-\mu_i)\ ,$$

and $r_i = (y_i - \mu_i)/\sqrt{V_i}$ are the Pearson's residuals. Note that when $k \to \infty$ and $w(x_i) = 1$, the right hand side of (16) becomes the score function for (15), giving the classical MLE as a special case of (16). When $k < \infty$ and $w(x_i) = 1$, we have a Huber-type estimator. One can fix k in order to have a pre-specified upper bound for the IF for a given (reasonable) level of contamination. The correction term $a(\mu_i)$ is included in order to ensure Fisher consistency. Note further that (16) can be seen as a direct generalization of robust approaches to regression and scale models.

From the general theory on M-estimation it follows that the estimator $\widehat{\beta}$, defined as the solution of (16), has bounded IF. The effect of outlying values in the y-direction is bounded by a finite value of the tuning constant k, and the effect of outlying values in x-direction is bounded by a suitable choice of the weights $w(\cdot)$.

The weights $w(x_i)$, $i = 1, \ldots, n$, are used to downweight leverage points. The available choices already discussed apply also to logistic regression. We can use the leverage

$$w(x_i) = \sqrt{1 - h_{ii}}, \tag{17}$$

or the (when covariates are continuous) MCD as described in Section 3.1. Victoria-Feser (2002) gives other suggestions for mixed types of covariates.

Cantoni and Ronchetti (2001) derive also asymptotics for these estimators, allowing inference (i.e., tests and confidence intervals for the odds ratios) to be performed in the usual way. Formally, we have that $\sqrt{n}(\widehat{\beta} - \beta)$ converges in distribution to

a zero-centered Gaussian as the sample size increases. A detailed derivation of the asymptotic variance is given in (Cantoni and Ronchetti, 2001, Appendix B).

As it is intuitive, under no contamination, standard errors are somewhat inflated with respect to the classical MLE, so that a little loss in power is expected. On the other hand, under contamination, tests based on the MLE are not reliable, and significant relationships may be often masked (see the application below for an example). It is also possible to use tests based on likelihood ratio test statistics: Adimari and Ventura (2001) show that an adjusted version of the quasi-profile likelihood ratio test statistic for a chosen scalar component of β has the standard asymptotic behaviour (i.e., it is chi-squared distributed with one degree of freedom). Along similar lines, Victoria-Feser (2002), building on results from Heritier and Ronchetti (1994), develops a robust version of Rao's score test, which follows standard asymptotic behaviour.

4.1 An example: Food Stamp Data

An example of binary logistic regression is proposed by Künsch *et al.* (1989). We have observed participation in the US Food Stamp Program, together with three covariates: tenancy, supplemental income, and monthly income. We have the sample size $n = 150$.

Table 3 shows the MLE for a logistic regression model, with standard errors and p-values.

Variable	Parameter Estimate	Standard Error	p-value
Intercept	0.93	1.62	0.5681
Tenancy	-1.85	0.53	0.0005
Suppl. Income	0.90	0.50	0.07365
log(Income+1)	-0.33	0.27	0.2228

Table 3: Summary of logistic regression model for the Food Stamp Data.

Only tenancy is significantly decreasing the odds of participating in the US Food Stamp Program, while the other covariates (and the intercept) are not significant.

A careful look at monthly income, nevertheless, shows a clear outlier in the x-direction. See for instance Figure 10, in which we plot Cook's distances. The outlying observation number 5 has no monthly income and no supplemental income, and does not participate in the US Food Stamp Program. Its fitted probability is equal to 0.716, indicating the model is predicting a participation in the program.

We underline that in other cases it could be not at all easy to identify malicious outliers, due to masking. In all cases, one can use a robust method. With the Cantoni and Ronchetti (2001) approach, using (17), we obtain results in Table 4.

The robust approach downweights 11 observations out of the 150, with $w(x_5) = 0.052$. Another possibly malicious observation, which was not detected by simple residual analysis on the non-robustly estimated model, is observation number 66. We have $y_{66} = 1$, but the fitted probability with the MLE is approximately

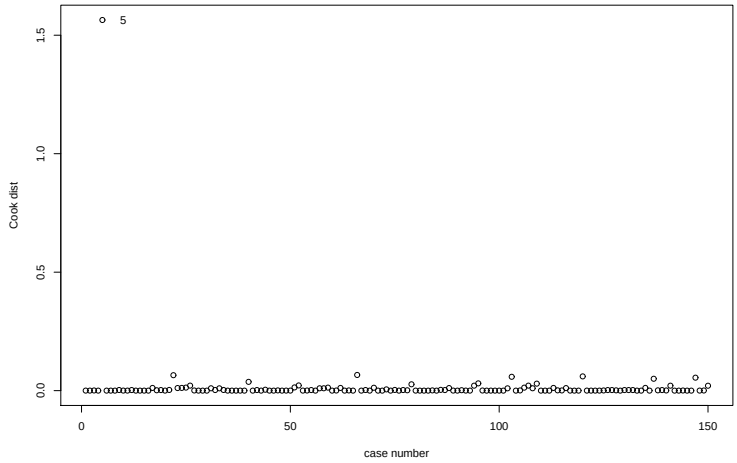

Figure 10: Cook Distances for the logistic regression model fit on Food Stamp Data.

Variable	Parameter Estimate	Standard Error	p-value
Intercept	6.49	3.03	0.0321
Tenancy	-1.83	0.58	0.0018
Suppl. Income	0.64	0.55	0.2428
log(Income+1)	-1.27	0.52	0.0144

Table 4: Summary of robust logistic regression model for the food stamp data.

$\Pr(Y_{66} = 1|x_{66}) = 0.04$. For this observation we have $w(x_{66}) = 0.154$. The other 9 downweighted observations all have weights above 0.45.

The robust estimates are quite different from the MLE. Most importantly, now also income has become significant, indicating that tenancy and high income both contribute to decreasing the odds of partiticipation in the food stamp program.

5 Robust survival analysis

Time to event data is often encountered in medical applications. These data are often analyzed by means of the semiparametric Cox model (Cox, 1972). Despite the unspecified baseline in the Cox model may be able to capture some aberrant behaviours, it can still happen that even a single malicious observation is unduly influent, with dramatic effects on parameter estimates. Outliers can lead to violation of the assumption of proportionality of hazard, and this departure may not be detected by common checking methods.

It has been well documented in the literature that the Cox model is sensitive even to slight departures from the assumptions (Samuels (1978), Bednarski (1989), Minder and Bednarski (1996)), and that its IF is not bounded (Reid and Crépeau,

1985). Valsecchi *et al.* (1996) provide a detailed illustration of how long survivors, for instance, may affect the estimates.

Many studies are devoted to diagnostics and assessing of robustness of the Cox model (e.g. Cain and Lange (1984)), but many of these proposals rely on residual analysis (Schoenfeld (1982), Grambsch and Therneau (1994), Therneau *et al.* (1990), Nardi and Schemper (1999)). Residuals are usually computed based on the non-robustly estimated model.

There are very few methods for robust estimation, which are described below. A more technical discussion can be found in (Heritier *et al.*, 2009, Ch. 7).

5.1 Robust estimation

Suppose we observe time to an event of interest for n independent subjects, and let (t_i, δ_i) denote the observed time and the event indicator for the i-th subject. Denote also by x_i a vector of subject specific covariates. In Cox proportional hazard regression (Cox, 1972) the effects of covariates on the hazard rate $\lambda(t|x_i)$ for the i-th individual is of the form

$$\lambda(t|x_i) = \lambda_0(t) \exp(\beta^\mathsf{T} x_i) \ , \quad i = 1, \ldots, n \ ,$$

where $\lambda_0(t)$ denotes a non-parametric baseline hazard. The hazard rate is easily linked to a subject specific survival distribution as

$$S(t|x_i) = \exp\left(- \int_0^t \lambda(u|x_i) \ du \right) .$$

The regression parameter β is estimated by maximizing the partial likelihood $L_P(\beta)$, where

$$L_P(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(\beta^\mathsf{T} x_i)}{\sum\limits_{t_j > t_i} \exp(\beta^\mathsf{T} x_j)} \right)^{\delta_i} . \tag{18}$$

The partial likelihood is estimated through an iterative Newton-Raphson algorithm, and the resulting maximum partial likelihood estimator (MPLE) is consistent and asymptotically normal under regularity conditions.

Methods for robust estimation are either based on weighting or trimming, which corresponds to giving zero weights to selected observations. Sasieni (1993a,b) uses a Wilcoxon-type weighting scheme. Outlying survival times are downweighted. Maximum likelihood proceeds like for the classical Cox regression, with the only difference that risk sets are weighted according to some criterion. The weights can be computed according to the number of subjects at risk (Breslow weights), to their square roots (Tarone-Ware weights) or according to the survival function estimates (Prentice weights). Downweighting risk sets with few subjects at risk or small survival estimates directly corresponds to downweighting long term survivors. The covariance matrix is estimated using a sandwich estimate (Lin, 1991); see also Sasieni (1993a). The approach is very flexible since many choices for the weighting method are available, and is particularly appealing since it provides an unbiased estimate

of the average hazard ration in case of non-proportional hazards (Schemper *et al.*, 2009).

Bednarski (1993), instead, smooths the partial likelihood by introducing weight functions inside the integral equation solved by the MPLE. This idea was refined in Bednarski (2007) to make it adaptive and invariant to time-transformation. There is a close connection between Bednarski and Sasieni methods, which is underlined in Bednarski and Nowak (2003).

Farcomeni and Viviani (2010) instead proceed by trimming the smallest contributions to the likelihood, which are more likely to be arising from outlying observations. Suppose there are $\lceil n(1-\alpha) \rceil$ clean observations, whose indices are collected in the set I^*, and that the remaining $\lfloor n\alpha \rfloor$ observations are instead outliers. Trimming is justified by the following contaminated model

$$\begin{cases} \lambda(t|x_i) = \lambda_0(t)\exp(\beta^\mathsf{T} x_i) & \text{if } i \in I^* \\ \lambda(t|x_i) = \lambda_i(t) & \text{if } i \notin I^*. \end{cases} \tag{19}$$

Contaminated observations arise from an unknown and observation-specific unspecified hazard rate $\lambda_i(t)$.

Denote with $H(\alpha)$ the set of all subsets of the vector of integers $(1, \ldots, n)$, where each of these subsets is of cardinality $\lceil n(1-\alpha) \rceil$. The MPLE for model (19) is the maximizer of

$$L_{TRIM}(\beta) = \max_{I \in H(\alpha)} \prod_{i \in I} \left(\frac{\exp(\beta^\mathsf{T} x_i)}{\sum\limits_{t_j > t_i, j \in I} \exp(\beta^\mathsf{T} x_j)} \right)^{\delta_i}. \tag{20}$$

That is, $\widehat{\beta}$ is the largest maximum over all possible maxima of the partial likelihoods computed only on subsets of $\lceil n(1-\alpha) \rceil$ observations.

In practice, α is not known, and the user will set the trimming level slightly larger than the expected proportion of contaminated observations. Based on the general work by Chakraborty and Chaudhury (2008), Farcomeni and Viviani (2010) propose a Metropolis-type algorithm, maximizing the trimmed partial likelihood for fixed α. A nonparametric bootstrap (Davison and Hinkley, 2006) is performed in order to estimate standard errors and confidence intervals.

A deep comparison between Bednarski and Sasieni methods can be found in Bednarski and Nowak (2003), while Farcomeni and Viviani (2010) compare the three robust methods and classical Cox regression with a brief simulation study. In all cases, classical Cox regression is seen to break down under contamination.

5.2 Outlier detection

Outliers in survival studies are interpreted by Nardi and Schemper (1999) as individuals whose failure time is too short, or too long, with respect to the median survival as predicted by the model. This definition is quite general and suggested Nardi and Schemper (1999) a clever method for identifying outliers. In addition to their work, we stress that, due to likelihood of masking residuals should always be

computed from the robustly estimated model. An illustration of this will be given below in the prostate cancer example.

For subjects experiencing the event, log-odds residuals are defined as

$$w_i = \log\left(\frac{\widehat{S}(t_i)}{2-\delta_i-\widehat{S}(t_i)}\right) \ , \quad i=1,\ldots,n \ ,$$

where $\widehat{S}(\cdot)$ is the estimated survival function. Under the null hypothesis of no contamination for the i-th subject, w_i asymptotically follows a standard logistic distribution and hence can be easily used for formally testing if the i-th observation is outlying. More details can be found in Nardi and Schemper (1999).

5.3 An example: Prostate cancer data

Data come from Andrews and Herzberg (1985) and was used by Nardi and Schemper (1999) to illustrate outlier detection in the Cox model.

Survival times were recorded for $n = 297$ patients with prostate cancer, together with seven binary prognostic factors: treatment, performance status (PS), serum Hemoglobin level in g/100 ml ($>$ 12-$\leq$ 12), weight index, history of cardiovascular disease, tumor size (Small-Large), and a combined index of tumor stage and grade.

The Cox model fit to the full data set gave estimates as shown in Table 5.

Variable	Parameter Estimate	Standard Error	p-value	Hazard Ratio	95% Confidence Limits
History	0.51	0.15	0.0005	1.66	1.251 - 2.215
Size	0.78	0.21	0.0002	2.19	1.453 - 3.299
Grade	0.69	0.15	¡0.0001	2.00	1.479 - 2.708
Weight	-0.33	0.15	0.0293	0.72	0.538 - 0.968
Hemoglobin	-0.25	0.18	0.1805	0.78	0.545 - 1.121
PS	0.1405	0.25	0.57	1.15	0.706 - 1.187
Treatment	0.05	0.17	0.7572	1.05	0.758 - 1.463

Table 5: Summary of Cox model for the Prostate Cancer Data.

Through the computation of log-odds residuals, four patients are flagged as outliers at level 0.05. Of these, two are censored patients with quite large survival times.

By applying the three methods for robust Cox regression, we obtain estimates and standard errors as reported in Table 6. The trimmed estimates, and subsequent outlier identification, are fairly stable with respect to the choice of α.

The trimmed and Bednarski methods identify the same four outliers as before, plus an additional two which were masked at non-robust estimation. Sasieni method identifies the same four outliers of the non-robust Cox model, but only one of the two additional outliers identified by the other methods.

From Table 6 it can be appreciated that robustly estimated hazard ratios, at least for the significant covariates, are generally slightly more extreme than the hazard ratios estimated by the classical Cox model. Consequently, the effect of risk

Variable	Trimmed Regression	Bednarski	Sasieni
History	0.55*	0.71*	0.49*
Size	1.00*	0.64*	0.69*
Grade	0.88*	0.85*	0.73*
Weight	-0.39*	-0.29	-0.33*
Hemoglobin	-0.31	-0.43*	-0.29
PS	-0.06	0.36	0.20
Treatment	0.09	0.06	0.049

Table 6: Summary of robust estimates for the Prostate Cancer Data. The trimming level is set as $\alpha = 0.1$, Sasieni method is based on Breslow weights. An asterisk indicates significance at the 5% level.

factors (like size and grade) may be underestimated by Cox model, resulting in overly optimistic survival prognosis and risk assessment for prostatic cancer patients.

6 Robust estimation of the area under the ROC curve

ROC curves are widely used to examine the effectiveness of continuous diagnostic markers in distinguishing between diseased and non-diseased individuals. ROC curves can be obtained under the assumption that the measurements of the diagnostic marker on the diseased and non-diseased subjects are distributed as two random variables X_1 and X_2, respectively. The area under the ROC curve (AUC) is the most popular summary measure of diagnostic accuracy of a continuous-scale test, or equivalently, of the diagnostic effectiveness of a continuous diagnostic marker. Its advantage consists of providing a single index that summarizes the overall performance of a diagnostic test or continuous marker, other than an entire curve. Values of the AUC close to 1 indicate very high diagnostic accuracy, while very low accuracy corresponds to values close to 0.5. Bamber (1975) showed that the AUC is equal to

$$A = P(X_1 < X_2)\,, \tag{21}$$

which can be interpreted as the probability that, in a randomly selected pair of diseased and non-diseased subjects, the diagnostic test value is higher for the diseased patient. In more general contexts, the AUC is also used as a measure of difference between distributions Wolfe and Hogg (1971).

ROC curves and the AUC have been studied under both parametric and nonparametric assumptions. There is a substantial literature on statistical inference for A under various parametric assumptions for X_1 and X_2; see, e.g., Kotz *et al.* (2003) and Pepe (2003). Parametric inference has been broadly handled by likelihood based procedures, theory of unbiased estimation or under a Bayesian perspective. Furthermore, some contributions addressing inference about A have also been provided in semiparametric and nonparametric settings; see, among others, the recent papers of Adimari and Chiogna (2006) and Qin and Zhou (2006).

Parametric inference about A assume that X_1 and X_2 are independent random variables with distribution functions $F_1 = F_1(x;\theta_1)$ and $F_2 = F_2(x;\theta_2)$, respectively.

Then, $A = A(\theta) = \int F_1(t;\theta_1)\, dF_2(t;\theta_2)$, with $\theta = (\theta_1, \theta_2)$. In this situation, although classical likelihood based procedures for inference on A are available, that they can be badly affected by mild departures from model assumptions, regarding both X_1 and X_2. To overcome this drawback, Greco and Ventura (2010) propose a robust inferential procedure for inference about A, based on the theory of bounded influence M-estimators for θ and the related tests. In particular, given a bounded influence M-estimator $\widehat{\theta}$ for $\theta = (\theta_1, \theta_2)$, the estimator $\widehat{A} = A(\widehat{\theta})$ is a bounded influence estimator for the AUC. Large-sample tests for robust inference on A are obtained by applying the delta-method and thus robust inference on A can be based on the studentuzed statistic

$$t(A) = \frac{(\widehat{A} - A)}{\sigma_A^2} ,$$

where σ_A^2 is a consistent estimator for the asymptotic variance of $\widehat{A}$ (see Greco and Ventura (2010)).

6.1 An example: ALCL lymphoma

The aim of this study was to assess the role of the Hsp70 protein in association with the anaplastic large cell (ALCL) lymphoma, which is a rare cancer disease which affects both children and adults. Diseased patients seem to have higher Hsp70 levels than healthy subjects (Mayer and Bukau, 2005). Moreover, excessive Hsp70 protein levels in diseased patients seem to limit the efficacy of the chemotherapy treatment. Thus, Hsp70 protein levels can be studied as a biomarker for detecting early ALCL lymphoma and therefore, its effectiveness in diagnosing the disease was evaluated by the AUC approach. The interest was also to interpret the AUC as the probability that the Hsp70 protein level is higher in ALCL cancer patients than in healthy individuals.

The data consist of a small sample: 10 patients with ALCL lymphoma in the group of cases and 4 healthy subjects in the group of controls. Hsp70 protein level was recorded on a continuous scale for each individual (see Cortese and Ventura (2009)). Two independent exponential random variables, $X_1 \sim \exp(\alpha)$ and $X_2 \sim \exp(\alpha)$, were assumed for the protein level in cancer patients and in non-diseased subjects, respectively. The two protein level samples result to have both different means (equal to 0.23 and 1.44 in the controls and cases, respectively) and variances (equal to 0.15 and 1.55 in the controls and cases, respectively), as observed in Figure 11. The values for the two OBREs (see Section 2.2) for the mean of the exponential distribution are equal to 0.27 and 1.37 in the controls and cases, respectively. The OBRE for the scale parameter θ of the exponential model is defined by the estimating function $\psi(y;\theta) = (a - \theta y)w$, with $w = \min(1, b/|a - \theta y|)$, for appropriate constants a and b. Here, we set $a = 0.86$ and $b = 1.13$ to get 95% efficiency at the exponential model (see Greco and Ventura (2010) for details).

In this framework, the AUC can be written as $A = \alpha/(\beta + \alpha)$. The MLEs for the exponential parameters, $\alpha_{mle} = 4.25$ and $\beta_{mle} = 0.70$, are substantially different in the two samples, suggesting thus a high value of the AUC, that is $A_{mle} = 0.86$. Confidence intervals (CI) for the AUC based on the classical Wald and robust Wald-type statistics (based on the OBRE) are reported in Table 7, together with the MLE

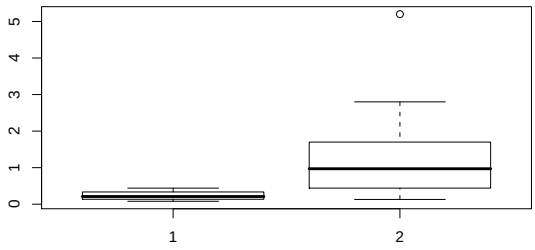

Figure 11: Boxplot of the Hsp70 protein level in controls (1) and cases (2).

AUC	*Point estimates (se)*	*Confidence intervals*
MLE	0.86 (0.07)	(0.719, 0.999)
MLE(-8)	0.81 (0.09)	(0.686, 0.896)
OBRE	0.83 (0.09)	(0.681, 0.922)

Table 7: Point estimates and 95% confidence intervals for the AUC in the ALCL lymphoma data.

and the robust point estimate for the AUC. We also estimated the AUC by means of the MLE method without the extreme observation on the sample of the second group (MLE(-8)). Table 7 reports that the estimated probaility that a cancer patient has higher Hsp70 protein level than a healthy patient is about 0.86 when MLE methods are used, and is about 0.83 when robust procedures are used. Moreover, robust CI seems to be more protective in estimating the accuracy of the protein level biomarker. Without the extreme observation in case subjects, the MLE provides an estimate similar to the robust estimator.

7 R functions for robust analyses

Basic procedures.
The simple function `mean` can be used to compute also a trimmed mean, using `mean(x,trim=0)`. The median and the MAD are can be computed using the `median` and the `mad` functions. The `huber` and `hubers` functions of the library `MASS` find the Huber's point estimators for scale and location parameters. To obtain the asymptotic variance of the estimators, it could be preferable to use the `rlm` function. Moreover, this function also gives the Tukey's bisquare and the Hampel's proposals. Also the robust versions of the classical t-test and the ANOVA model can be performed with the `rlm` function.

Linear regression.
The simultaneous estimation of β and σ is obtained using iteratively reweighted least squares, in which both estimators are updated at each iteration. Robust M-

estimation of scale and regression parameters can be performed using the `rlm` function in `R`. The choice `psi.huber` with $k = 1.345$ for β and MAD of the residuals for σ represents the default in `rlm`. The choice `psi="psi.bisquare"` in `rlm` gives the Tukey's redescending estimator.

The `R` function `wle.lm` allows to fit a linear model with the weighted likelihood, when the errors are independent and identically distributed random variables from a Gaussian distribution.

The function `rlm` has an option that allows to implement MM-estimation, that is `method="MM"`. To use other breakdown point estimates, in `R` there exists the function `lqs` to fit a regression model using resistant procedures, that is achieving a regression estimator with a high breakdown point (see Rousseeuw and Leroy, 1987, Marazzi, 1993, and Venables and Ripley, 2002, Sec. 6.5).

Robust Logistic Regression.
The functions `glmrob` in the package `robustbase` and `glmRob` in package `robust` allow to fit robust logistic regression models, with different choices for the weight functions.

Survival Analysis.
R code for Sasieni approach can be found in function `coxphw` in the package `coxphw`. The user needs to choose the type of weighting scheme used.

R code for Bednarski approach can be found in function `coxr` in the package `coxrobust`.

R code for the Farcomeni and Viviani (2010) approach is freely available from the Web page `http://afarcome.interfree.it/robcox.r`.

References

G. Adimari and M. Chiogna (2006). Partially parametric interval estimation of Pr(Y>X). *Computational Statistics and Data Analysis*, **51**, 1875–1891.

G. Adimari and L. Ventura (2001). Robust inference for generalized linear models with application to logistic regression. *Statistics & Probability Letters*, **55**, 413–419.

G. Adimari and L. Ventura (2002a). Quasi-likelihood from M-estimators: A numerical comparison with empirical likelihood. *Statistical Methods and Applications*, **11**, 175–185.

G. Adimari and L. Ventura (2002b). Quasi-profile loglikelihoods for unbiased estimating functions. *Annals of the Institute of Statistical Mathematics*, **54**, 235–244.

C. Agostinelli and M. Markatou (1998). A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. *Statistics and Probability Letters*, **37**, 341–350.

D.F. Andrews and A.M. Herzberg (1985). *Data : a collection of problems from many fields for the student and research worker*. Springer-Verlag, New York.

A.C. Atkinson and M. Riani (2000). *Robust diagnostic regression analysis*. Springer-Verlag, New York.

D. Bamber (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Mathematical Psycology*, **12**, 387–415.

T. Bednarski (1989). On sensitivity of Cox's estimator. *Statistics and Decisions*, **7**, 215–228.

T. Bednarski (1993). Robust estimation in Cox regression model. *Scandinavian Journal of Statistics*, **20**, 213–225.

T. Bednarski (2007). On a robust modification of Breslow's cumulated hazard estimator. *Computational Statistics and Data Analysis*, **52**, 234–238.

T. Bednarski and M. Nowak (2003). Robustness and efficiency of Sasieni-type estimators in the Cox model. *Journal of Statistical Planning and Inference*, **115**, 261–272.

R. Bellio (2007). Algorithms for bounded-influence estimation. *Computational Statistics and Data Analysis*, **51**, 2531–2541.

R. Bellio, L. Greco, and L. Ventura (2008). Adjusted quasi-profile likelihoods from estimating functions. *Journal of Statistical Planning and Inference*, **138**, 3059–3068.

A. Bianco, G. Boente, and J. di Rienzo (2000). Some results for robust GM-based estimators in heteroscedastic regression models. *Journal of Statistical Planning and Inference*, **89**, 215–242.

A.M. Bianco and V.J. Yohai (1997). Robust estimation in the logistic regression model. In: H. Rieder, ed., *Robust Statistics, Data Analysis and Computer Intensive Methods*, 17–34. Springer, New York.

A.R. Brazzale, A.C. Davison, and N. Reid (2007). *Applied Asymptotics*. Cambridge University Press, Cambridge.

K.C. Cain and N.T. Lange (1984). Approximate case influence for the proportional hazards regression model with censored data. *Biometrics*, **40**, 493–499.

E. Cantoni and E. Ronchetti (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, **96**, 1022–1030.

R.J. Carroll and S. Pederson (1993). On robustness in the logistic regression model. *Journal of the Royal Statistical Society, Series B*, **55**, 693–706.

R.J. Carroll and D. Ruppert (1982). Robust estimation in heteroscedastic linear models. *The Annals of Statistics*, **10**, 429–441.

R.J. Carroll and D. Ruppert (1988). *Transformation and weighting in regression*. Chapman and Hall, London.

B. Chakraborty and P. Chaudhury (2008). On an optimization problem in robust statistics. *Journal of Computational and Graphical Statistics*, **17**, 683–702.

R. D. Cook and S. Weisberg (2007). *Residuals and influence in regression*. Chapman and Hall, London.

J.B. Copas (1988). Binary regression models for contaminated data. *Journal of the Royal Statistical Society, Series B*, **50**, 225–265.

S. Copt and S. Heritier (2007). Robust alternatives to the F-test in mixed linear models based on MM-estimates. *Biometrics*, **63**, 1045–1052.

S. Copt and M.P. Victoria-Feser (2006). High-breakdown inference for mixed linear models. *Journal of the American Statistical Association*, **101**, 292–300.

G. Cortese and L. Ventura (2009). Accurate likelihood on the area under the ROC curve for small samples. *Tech. Rep. 17*, Department of Statistics, University of Padova.

D. R. Cox (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

J. Cuesta-Albertos, A. Gordaliza, and C. Matrán (1997). Trimmed k-means: an attempt to robustify quantizers. *Annals of Statistics*, **25**, 553–576.

A. C. Davison and D. Hinkley (2006). *Bootstrap Methods and their Applications*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge.

A. Farcomeni (2009). Robust Double Clustering: A Method Based On Alternating Concentration Steps. *Journal of Classification*, **26**, 77–101.

A. Farcomeni and S. Viviani (2010). Trimmed Cox regression for robust estimation in survival studies. *Tech. Rep. 7*, Department of Statistics, Sapienza - University of Rome.

M.T. Gallegos and G. Ritter (2005). A robust method for cluster analysis. *Annals of Statistics*, **33**, 347–380.

L.A. Garcia-Escudero and A. Gordaliza (1999). Robustness properties of k means and trimmed k means. *Journal of the American Statistical Association*, **94**, 956–969.

D.M. Giltinan, R.J. Carroll, and D. Ruppert (1986). Some new estimation methods for weighted regression when there are possible outliers. *Technometrics*, **28**, 219–230.

P. M. Grambsch and T. M. Therneau (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**, 515–526.

L. Greco and L. Ventura (2010). Robust inference for the stress-strength reliability. *Statistical Papers*, to appear.

F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

J.J. Hanfelt and K.Y. Liang (1995). Approximate likelihood ratios for general estimating functions. *Biometrika*, **82**, 461–477.

S. Heritier, E. Cantoni, S. Copt, and M.-P. Victoria-Feser (2009). *Robust methods in Biostatistics*. Wiley, Chichester, U.K.

S. Heritier and E. Ronchetti (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, **89**, 897–904.

C.C. Heyde (1997). *Quasi-Likelihood and its application*. Springer-Verlag, New York.

P. J. Huber (1964). Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**, 73–101.

P. J. Huber (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, **1**, 799–821.

P. J. Huber (1981). *Robust Statistics*. Wiley, New York.

M. Hubert, Rousseeuw P.J., and S. Van Aelst (2008). High-breakdown robust multivariate methods. *Statistical Science*, **23**, 92–119.

M. Hubert, P.J. Rousseeuw, and K. Vanden Branden (2005). ROBPCA: a new approach to robust principal components analysis. *Technometrics*, **47**, 64–79.

M. Hubert and K. Van Driessen (2004). Fast and robust discriminant analysis. *Computational statistics and data analysis*, **45**, 301–320.

L. Kaufman and P.J. Rousseeuw (1990). *Finding groups in data*. Wiley.

S. Kotz, Y. Lumelskii, and M. Pensky (2003). *The Stress-Strength Model and its Generalizations. Theory and Applications*. World Scientific, Singapore.

H.R. Künsch, L.A. Stefanski, and R.J. Carroll (1989). Conditionally unbiased bounded influence estimation in general regression models, with applications to generalized linear models. *Journal of the American Statistical Association*, **84**, 460–466.

G. Li (1985). Robust regression. *In: Exploring Data Tables, Trends, and Shapes, eds. Hoagling and Tukey*, 281–343.

D. Lin (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, **86**, 725–728.

A. Marazzi (1993). *Algorithms, Routines, and S Functions for Robust Statistics*. Wadsworth and Brooks/Cole, Pacific Grove.

M. Markatou, A. Basu, and B. Lindsay (1997). Weighted likelihood estimating equations: the discrete case with application to logistic regression. *Journal of Statistical Planning and Inference*, **57**, 215–232.

M. Markatou, A. Basu, and B.G. Lindsay (1998). Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, **93**, 740–750.

M. Markatou and T.P. Hettmansperger (1990a). Applications of the asymmetric eigenvalue problem techniques to robust testing. *Journal of Statistical Planning and Inference*, **31**, 51–65.

M. Markatou and T.P. Hettmansperger (1990b). Robust bounded-influence tests in linear models. *Journal of the American Statistical Association*, **85**, 187–190.

R. A. Maronna, R. D. Martin, and V. J. Yohai (2006). *Robust Statistics: Theory and Methods*. Wiley, New York.

R.A. Maronna (2005). Principal components and orthogonal regression based on robust scales. *Technometrics*, **47**, 264–273.

M.P. Mayer and B. Bukau (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, **62**, 670–84.

P. McCullagh (1991). Quasi-likelihood and estimating functions. In: *Statistical Theory and Modelling. In Honour of Sir David Cox.*, 265–286. Chapman and Hall.

P. McCullagh and J. A. Nelder (1989). *Generalized Linear Models, 2nd Edition*. Chapman and Hall, CRC, London.

J.W. McKean, S.J. Sheather, and T.P. Hettmansperger (1993). The use and interpretation of residuals based on robust estimation. *Journal of the American Statistical Association*, **88**, 1254–1263.

C.E. Minder and T. Bednarski (1996). A robust method for proportional hazards regression. *Statistics in Medicine*, **15**, 1033–1047.

S. Morgenthaler (1992). Least-absolute-deviations fits for generalized linear models. *Biometrika*, **79**, 747–754.

A. Nardi and M. Schemper (1999). New Residuals for Cox Regression and Their Application to Outlier Screening. *Biometrics*, **55**, 523–529.

A.B. Owen (2001). *Empirical likelihood*. Chapman and Hall, London.

M.S. Pepe (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, Oxford.

L. Piron, A. Turolla, M. Agostini, C.S. Zucconi, L. Ventura, P. Tonin, and M. Dam (2010). Motor learning principles for rehabilitation: A pilot randomized controlled study in poststroke patients. *Neurorehabilitation and Neural Repair*, in press.

D. Pregibon (1982). Resistant fits for some commonly used logistic models with medical applications. *Biometrics*, **38**, 485–498.

G.S. Qin and X.H. Zhou (2006). Empirical likelihood inference for the area under the roc curve. *Biometrics*, **62**, 613–622.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

N. Reid and H. Crépeau (1985). Influence function for proportional hazards regression. *Biometrika*, **72**, 1–9.

P.J. Rousseeuw (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871–880.

P.J. Rousseeuw and A.M. Leroy (1987). *Robust regression and outlier detection*. Wiley-Interscience, New York.

A.D. Rule, T.S. Larson, E.J. Bergstralh, J.M. Slezak, Jacobsen S.J., and F.G. Cosio (2004). Using serum creatinine to estimate glomerular filtration rate: Accuracy in good health and in chronic kidney disease. *Annals of Internal Medicine*, **141**, 929–938.

S. Samuels (1978). *Robustness for survival estimators*. Ph.D. thesis, Department of Biostatistics, University of Washington.

P.D. Sasieni (1993a). Maximum weighted partial likelihood estimates for the Cox model. *Journal of the American Statistical Association*, **88**, 144–152.

P.D. Sasieni (1993b). Some new estimators for Cox regression. *Annals of Statistics*, **21**, 1721–1759.

M. Schemper, S. Wakounig, and G. Heinze (2009). The estimation of average hazard ratios by weighted Cox regression. *Statistics in Medicine*, **28**, 2473–2489.

D. Schoenfeld (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, **69**, 239–241.

R.G. Staudte and S.J. Sheather (1990). *Robust Estimation and Testing*. Wiley, New York.

T. M. Therneau, P. M. Grambsch, and T. Fleming (1990). Martingale based residuals for survival models. *Biometrika*, **77**, 147–160.

M.G. Valsecchi, D. Silvestri, and P. Sasieni (1996). Evaluation of long-term survival: use of diagnostics and robust estimators with Cox's proportional hazards models. *Statistics in Medicine*, **15**, 2763–2780.

W. N. Venables and B. D. Ripley (2002). *Modern Applied Statistics with S*. Springer, New York.

M.-P. Victoria-Feser (2002). Robust inference with binary data. *Psychometrica*, **67**, 21–32.

R.W.M. Wedderburn (1974). Quasi-likelihood function, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–447.

A. Welsh and A. Richardson (1997). Approaches to the robust estimation of mixed models. *In: Handbook of Statistics, eds. Maddala G and Rao C*, **15**, 343–385.

A.H. Welsh and E. Ronchetti (1985). A journey in single steps: Robust one-step M-estimation in linear regression. *Journal of Statistical Planning and Inference*, **103**, 287–310.

D.A. Wolfe and R.V. Hogg (1971). On constructing statistics and reporting data. *The American Statistician*, **25**, 27–30.

V.J. Yohai (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, **15**, 642–656.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy