An overview of robust methods in medical research

Alessio Farcomeni and Laura Ventura

Stat Methods Med Res published online 25 October 2010
DOI: 10.1177/0962280210385865

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–23

DOI: 10.1177/0962280210385865
smm.sagepub.com
SAGE

An overview of robust methods in medical research

Alessio Farcomeni[1] and Laura Ventura[2]

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 article, we review and outline how robust inferential procedures may routinely be applied in practice in the biomedical 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. Under this assumption, parametric inferential procedures based on the sample mean, standard deviation, one- and two-samples *t*-test, and so on, are the most efficient. 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. In this case, it may be preferable to base inference on procedures that are more resistant, that is which specifically take into account the fact that the assumed models used by the analysts are only approximate. 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.[1–4]

Typically, biostatisticians continue to use *ad hoc* techniques to deal with outliers and underestimate the impact of model misspecifications. However, removing outliers with a simple data screening and then applying classical inferential procedures to remaining data can be flawed in many respects. First of all, in multivariate or highly structured settings, outliers can be masked and very hard to detect. Second, in many cases it is more efficient to down-weight instead of discard these observations.

[1]Department of Public Health and Infectious Diseases, Sapienza University of Rome, Rome, Italy.
[2]Department of Statistics, University of Padua, Padua, Italy.

Corresponding author:
Alessio Farcomeni, Department of Public Health, Sapienza University of Rome, Rome, Italy.
Email: alessio.farcomeni@uniroma1.it

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. The aim of this article 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 article we overview and illustrate the use of some robust procedures for a variety of data structures which statisticians often encounter in practice, such as the well known one- and two-sample t-tests, regression models, logistic regression, survival analysis and receiver operating characteristic (ROC) curves. We mainly focus on robustness with respect to outliers.

A related account may be found in Heritier et al.,[4] 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 principal components analysis,[5–7] clustering,[8–11] double clustering,[12] discriminant analysis[13] and mixed linear models.[14] Our account is far from being complete, and we point the reader to the several books on robust statistics, see, for example, Heritier *et al*.[4] and Rousseeuw and Leroy,[5] and the more theoretical Huber,[1] Hampel *et al*.[2] and Maronna *et al*.[3]

The outline of this article 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 ROC curves 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

The sample mean, and thus also the classical t-test or the analysis of variance (ANOVA) model, can be badly affected by mild departures from model assumptions. Typically, when the normality assumption is questionable, non-parametric methods are used. We note that non-parametric methods simply use different (often, more general) assumptions, for instance assuming that the treatment only shifts the unknown distribution, whose shape is instead not affected. Robust methods can be a better choice since they can be calibrated to have a small loss of efficiency with respect to parametric tests, consequently being more powerful than rank-based methods, and they are by construction resistant to some violations of assumptions.

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- 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.[2]

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 $F(y; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}^p$, $p \geq 1$. By slight model misspecification, we mean that the data-generating process lies in a neighbourhood of F_θ, that is considered a sensible model for the problem under investigation.

The notion of neighbourhood can be formalised as $F_\varepsilon = (1-\varepsilon)\ F_\theta + \varepsilon G$, where G is an arbitrary distribution and $0 \leq \varepsilon \leq 1$.[15]

One way of assessing the robustness of a statistic $T = T(y)$ is by means of the influence function (IF):

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

where $T(F_{\varepsilon,\Delta}) = T((1-\varepsilon)F_\theta + \varepsilon\Delta_x)$, with Δ_x mass probability at the point x, that is, $\Pr(\Delta_x = x) = 1$. The IF measures the local stability of a statistical procedure, since it describes the effect of an infinitesimal contamination at the point x. The supremum of the IF, that is the gross-error sensitivity (GES), measures the worst influence on T and a desiderable robustness property is a finite GES, that is 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.[2]

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 it breaks down. 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 centre of the data such that points lying outside this distance have no influence on the asymptotic bias. Formally, for a symmetric F_θ centred at m, RP is defined as $inf\{r > 0 : \mathrm{IF}(x;\ T,\ F_\theta) = 0 \text{ when } \delta(x, m) > r\}$, where δ is a suitable distance measure. If an estimator has a finite RP, then points too far away from the centre 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. An M-estimator of θ is implicitly defined as a solution $\widehat{\theta}$ of the unbiased estimating equation:[2]

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

where $\psi(\cdot)$ is a suitable function. If $\psi(\cdot)$ is the score function, then the maximum likelihood estimator (MLE) is obtained (e.g. the mean). Robust estimators are obtained with an opportune choice for $\psi(\cdot)$. For instance, 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 is consistent and approximately normal with mean θ and variance $V(\theta) = B(\theta)^{-1}\ \Omega(\theta)(B(\theta)^{-1})^{\mathrm{T}}$, where $B(\theta) = -E_\theta(\partial\Psi_\theta/\partial\theta^{\mathrm{T}})$ and $\Omega(\theta) = E_\theta(\Psi_\theta\Psi_\theta^T)$. Thus, confidence intervals (CI) (or testing of 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:

$$(\widehat{\theta}-\theta)^{T} V(\widehat{\theta})^{-1}(\widehat{\theta}-\theta), \tag{3}$$

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

$$t(\theta)=\frac{(\widehat{\theta}-\theta)}{V(\widehat{\theta})^{1/2}} \tag{4}$$

may be used to set CI or compute a p-value for θ.

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

$$\mathrm{IF}(x;\widehat{\theta},F_\theta)=B(\theta)^{-1}\psi(x;\theta)\ . \tag{5}$$

In view of (5), any bounded $\psi(x;\theta)$ yields a B-robust M-estimator, and a robust test statistic of the form (3) or (4).[16] 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, such as the well known Huber's ψ-function for location given by $\psi(y;\theta)=\psi_H(y-\theta;k)$, with:

$$\psi_H(x;k)=\max(-k,\min(k,x))\ , \tag{6}$$

where k is a user-specified constant. 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, that is 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%. The value $k=1.345$ gives 95% efficiency at the normal model. The efficiency of Huber's estimators is increasing with increasing k, and the robustness is decreasing with incrasing k. For the scale parameter Huber's Proposal 2 estimate is often used, with ψ-function given by $\psi_H(y;k)^2-k_1$, with k_1 being a known constant. Alternative estimators are given by the ψ-function of the optimal B-robust estimator (OBRE),[2] and by the so-called redescending 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.[1,2,17]

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.[18] 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- and pre-treatment measurements) in the two groups of patients. 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

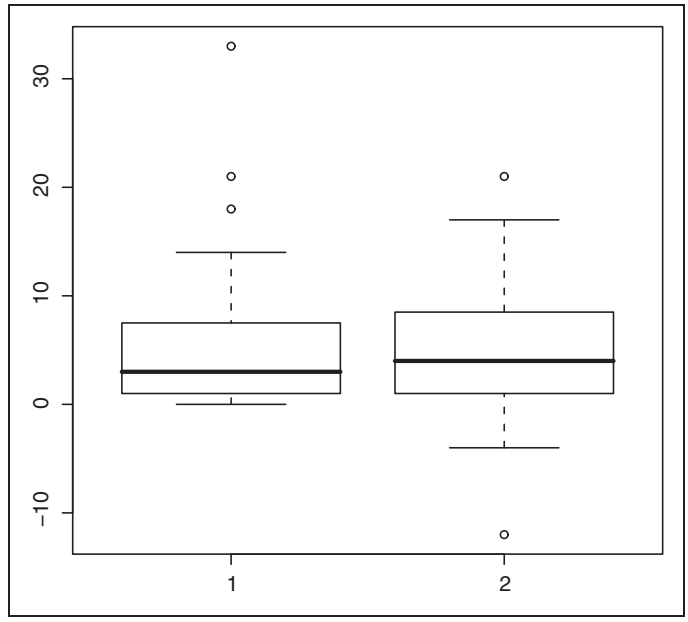

Figure 1. Boxplot of the differences pre-post FIM scores in case (1) and control patients (2).

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

	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)	

robust estimators are considered, the Huber's estimator is larger in the second group (as well as the median). 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 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.

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- and pre- 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

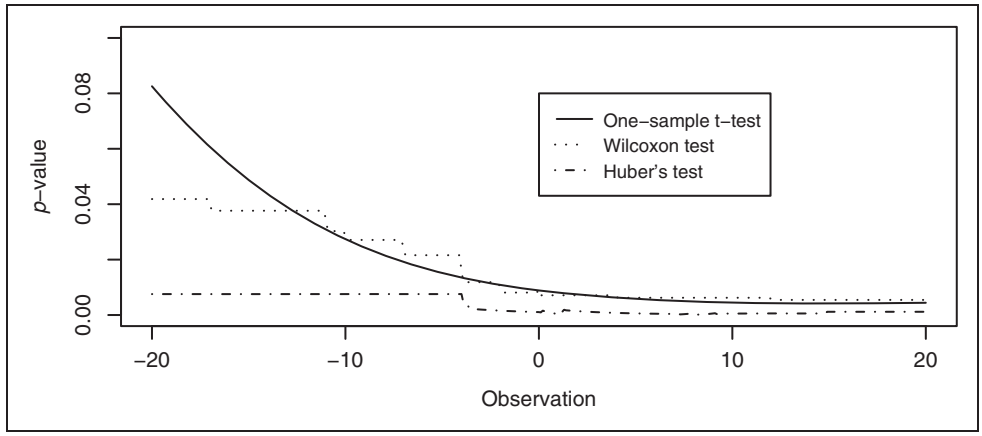

Figure 2. Sensitivity analysis of the *p*-values for the *t*-test, Wilcoxon test and Huber's test when an observation in differences pres-post FIM scores in control patients is changed.

the *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 the observation is greater than -2.

3 Robust inference in scale and regression models

The aim of this section is to give an overview of robust procedures to fit, test and check a scale and regression model. The main references for the arguments discussed here are, among others, Huber,[1] Hampel et al.,[2] Heritier et al.,[4] Venables and Ripley,[17] Li,[19] and references therein. A scale and regression model is given by:

$$y_i = x_i^T \beta + \sigma \varepsilon_i \ , \quad i = 1, \ldots, n \ , \tag{7}$$

where y_i is the response, x_i is a known *p*-variate vector of regressors, $\beta \in \mathrm{IR}^p$ $(p > 1)$ is an unknown 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 $p_0(x) = \phi(x)$, with $\phi(x)$ standard normal density, and exact inference based on the *t* and *F* distributions may then be performed using classical least squares estimates (LSE) or MLEs, available in any statistical package. More generally, models of form (7) are characterised by a linear predictor and possibly non-normal errors. Popular choices for $p_0(\cdot)$ include the Student's *t*, extreme value, logistic, skew-normal and Cauchy distributions (see, e.g.[20]).

Classical estimates for the parameters of model (7) are optimal, but they can be badly affected by mild departures from the assumed model. There are two possible sources of errors: the observations in the response and the *p*-variate vector of regressors (see Figure 3 for simple regression), that is, outliers in the *y*- and in the *x*-directions. In some situations of practical interest errors in the regressors can be ignored (such as in most bioassay experiments). Unusual observations in

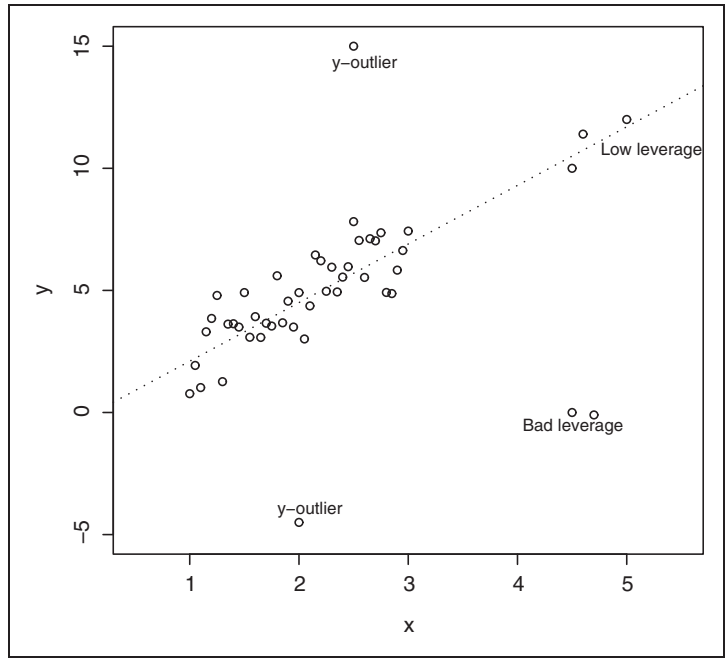

Figure 3. Different types of outliers in simple regression models.

x-direction are often 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, or vice versa. These kind of points can be very dangerous since they are typically very influential.

Outliers in y-direction are dealt with Huber-type estimators. Bounded influence estimators may be preferable with a moderate contamination, while with strong contamination (up to 50% of unusual observations), we may use high BP estimators. 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[21]) distance or forward searches for outliers detection.[22] However, although apparently simple, these strategies can be not only impractical, but also misleading; see for instance Welsh and Ronchetti.[23]

In this section, some theoretical robust methods in linear regression are inevitably skipped. For instance, situations with heteroscedastic models[24–26] and models for repeated measures.[14,27,28]

3.1 *M*-estimation

A wide class of robust M-estimators for the regression coefficients is defined by estimating functions of the form:

$$\Psi_\beta = \sum_{i=1}^{n} s(x_i)\psi_\beta(r_i) = \sum_{i=1}^{n} w_\beta(r_i) r_i x_i \ , \tag{8}$$

where $r_i = (y_i - x_i^T \beta)/\sigma$ is 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 Hample *et al.*,[2] 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, $w_\beta(r)=w_\beta^H(r)=\psi_H(r;k)/r=\min(1,k/|r|)$. The weights $w_\beta(r_i)$, $i=1,\ldots,n$, are interpretable and they automatically define a measure of outlyingness. An M-estimating equation is used for the scale parameter σ, that is $\Psi_\sigma=\sum_{i=1}^n \psi_\sigma(r_i)=0$, 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, 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 $0\le s(x)\le 1$ and $\psi_\beta(\cdot)=\psi_H(\cdot;\ k)$ in (8).[2,3] 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:

$$d_i=\sqrt{(x_i-\widehat{\mu})^T\widehat{\Sigma}^{-1}(x_i-\widehat{\mu})}\ . \tag{9}$$

In (9), $(\widehat{\mu},\widehat{\Sigma})$ are robust estimates of multivariate location and scatter, such as the minimum covariance determinant (MCD) estimator of Rousseeuw;[29] see also Hubert *et al.*[30] The weights can thus be defined as $s(x_i)=1$ if $d_i^2\le\chi^2_{p,0.975}$, and $s(x_i)=0$ otherwise. A further possibility to choose a weighting scheme in (8), which can be used also for categorical regressors, is based on the leverages h_{ii}, that is on the diagonal elements of $H=X(X^{\mathrm{T}}\ X)^{-\mathrm{T}}\ X^{\mathrm{T}}$, which are well known in regression diagnostic.[21]

The Hampel–Krasker estimator is optimal for $p_0(x)=\phi(x)$ in the sense that it is the M-estimator which minimises the trace of the asymptotic covariance matrix under the constraint that it has a bounded IF. In the context of non-normal linear regression models, Carroll and Ruppert[31] discuss general procedures to obtain the OBRE. To compute the OBRE an iterative algorithm is required.[32]

3.2 Other robust estimators

The IF is not the only important robustness measure. Another key concept is the BP (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. However, redescending estimating functions may have multiple roots, and this may considerably complicate the computation of the estimates.

Several alternative high BP and computationally efficient estimators of regression have been proposed, such as the least median of squares (LMS) or the least trimmed squares (LTS) estimate. The LMS minimises the median of the residuals, while the LTS minimises the trimmed mean of the residuals. 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[33,34] and have an asymptotic efficiency as close to one as desired, and simultaneously BP 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 chosen.

A final approach we mention is based on the weighted likelihood,[35,36] which builds a weight function $w(\cdot)$ depending on data y and on the distribution of the assumed parametric model.

Robust estimators 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.

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 (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.[16] These robust likelihoods can be used to define robust likelihood ratio-type tests. In particular, starting from an unbiased estimating function for β, quasi-likelihoods for β can be derived following Mccullagh,[37] Adimari and Ventura[38] and Bellio *et al.*,[39] and, see, also, Heritier and Ronchetti,[16] Markatou and Hettmansperger[40,41] and Hanfelt and Liang.[42] An alternative approach considers the empirical likelihood ratio statistic.[43]

For applications of quasi- and empirical likelihoods for robust inference in linear models see, among others, Markatou and Hettmansperger,[40,41] Adimari and Ventura,[38,44] Bellio *et al.*,[39] and Heritier et al.[4] (Section 3.3.3).

3.4 Model checking

As in standard regression analyses, it may be important to check the model assumptions, evaluating for instance the approximate normality of the estimated standardised 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.[45] 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,[4] given by:

$$R_r^2 = \left(\frac{\sum_{i=1}^{n} w_i(y_i - \bar{y}_w)(\widehat{y}_i - \bar{\widehat{y}}_w)}{\sqrt{\sum_{i=1}^{n} w_i(y_i - \bar{y}_w)^2 \sum_{i=1}^{n} w_i(\widehat{y}_i - \bar{\widehat{y}}_w)^2}} \right)^2, \tag{10}$$

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

3.5 Example: glomerular filtration rate data

This dataset, discussed also in Heritier *et al.*,[4] contains measurements of the glomerular filtration rate (`gfr`) and serum creatinine (`cr`) on $n = 30$ subjects. 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, synthsised 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). Note that there are some differences between robust estimates and the LSE. In particular, 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 BP estimators behave quite similarly.

Figure 5 gives possible diagnostic plots: the Cook's distance plot, the Cook's distance statistic versus $h_{ii}/(1-h_{ii})$ and the weights of the robust Tukey's fit. 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). Note also 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.

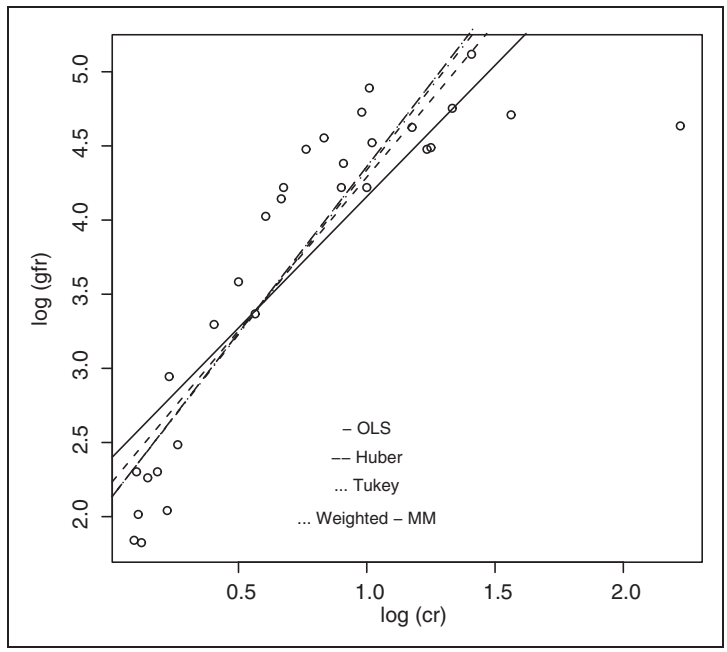

Figure 4. `GFR data`: LSE fit and several robust fits.

It may be useful to compare the different weights in the robust estimations visualising them (Figure 6). 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.*,[46] 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 corresponding p-values for significance testing based on Wald statistics. The conclusions of the fits

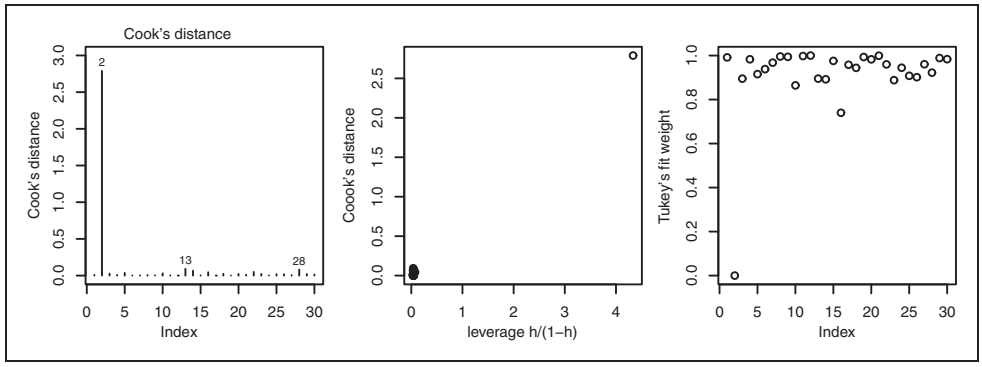

Figure 5. GFR data: diagnostic plots.

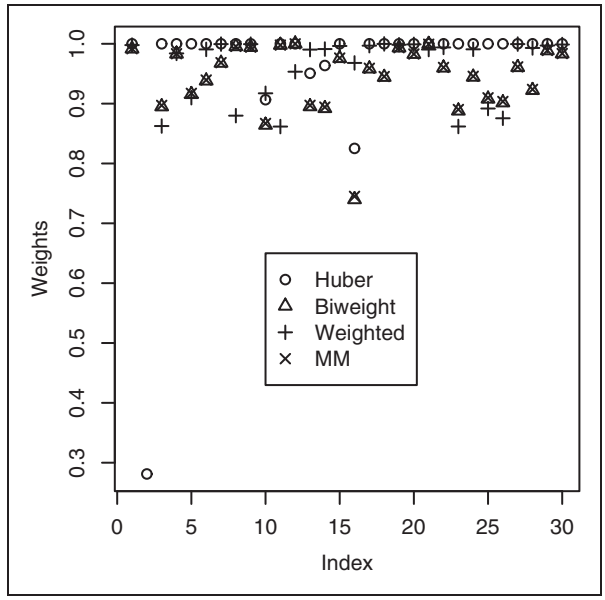

Figure 6. GFR data: weights of different robust estimates.

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

	LSE estimate (se)	*p*-value	Huber estimate (se)	*p*-value	Weighted estimate (se)	*p*-value	*MM* 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
σ	0.27		0.33		0.25		0.29	
R_r^2	0.943		0.943		0.952		0.88	

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 give different results on both the estimated values of the regression coefficients and in terms of R_r^2.

4 Robust logistic regression

Departures from model assumptions and malicious observations can lead to problems also in generalised 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:

$$\log\left(\frac{\Pr(Y_i = 1|x_i)}{1 - \Pr(Y_i = 1|x_i)}\right) = x_i^T \beta, \tag{11}$$

where we have used the popular *logit* link. See, for instance, McCullagh and Nelder[47] 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.[48–54] Note that there are also methods derived for robust estimation in the entire GLM class, as for instance the OBRE by Künsch *et al.*[55]

Here, we review a popular approach due to Cantoni and Ronchetti,[56] who develop a Mallows-type estimator based on a modification of the system of estimating equations derived from the quasi-likelihood estimator.[57,58] This latter estimator is the solution of the system of estimating equations:

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

where $\mu_i = e^{x_i^T \beta}/(1 + e^{x_i^T \beta})$, $V_i = \mu_i(1 - \mu_i)$, $i = 1, \ldots, n$. The Cantoni and Ronchetti[56] approach uses an appropriate weighting scheme and the Huber's $\psi_H(\cdot; k)$ function as follows:

$$\sum_{i=1}^{n} w(x_i) x_i^T (\psi_H(r_i; k) - a(\mu_i))\sqrt{V_i} = 0, \tag{12}$$

where

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

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 (12) becomes the score function for (11), giving the classical MLE as a special case of (12). 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 (12) can be seen as a direct generalisation of robust approaches to regression and scale models. The estimator $\widehat{\beta}$, defined as the solution of (12), 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 available choices already discussed apply also to logistic regression. We can use the leverage

$$w(x_i) = \sqrt{1 - h_{ii}}, \tag{13}$$

or the (when covariates are continuous) MCD as described in Section 3.1. Victoria-Feser[54] gives other suggestions for mixed types of covariates.

Cantoni and Ronchetti[56] derive also asymptotics for these estimators, allowing inference (i.e. tests and CI 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-centred Gaussian as the sample size increases. A detailed derivation of the asymptotic variance is given in Cantoni and Ronchetti,[56] 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[59] 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, building on results from Hertier and Ronchetti,[16] a robust version of Rao's score test following standard asymptotic behaviour can be developed.[54]

4.1 An example: food stamp data

We now illustrate robust logistic regression on data from Künsch *et al*;[55] see also Heritier and Ronchetti.[16] The response (y) indicates participation in the Federal Food Stamp programme and the covariates considered in the study include two dichotomous variables, tenancy (x_1) and supplemental income (x_2), and a logarithmic tranformation of monthly income [log(monthly income + 1)] (x_3). The US Food Stamp programme is designed to give assistance to families in need in the country, by distribution of cards which can be used to purchase food. We consider the model

$$\log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i},$$

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

	MLE estimate	Standard error	p-Value
β_0	0.93	1.62	0.5681
β_1	−1.85	0.53	0.0005
β_2	0.90	0.50	0.0736
β_3	−0.33	0.27	0.2228

with $i = 1, \ldots, 150$. Table 3 shows the MLE for the logistic regression model, with standard errors and p-values. Only tenancy is significantly decreasing the odds of participating in the US Food Stamp programme, while the other covariates (and the intercept) are not significant.

Previous analyses highlighted that the data contain at least a leverage point (Case 5); see also Figure 7, 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 programme. Its fitted probability is equal to 0.716, indicating the model is predicting a participation in the programme.

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[56] approach, using (13), 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 $\Pr(Y_{66} = 1|x_{66}) = 0.04$. 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 participation in the Food Stamp Programme.

5 Robust survival analysis

Time to event data is often encountered in medical applications. These data are often analysed by means of the semiparametric Cox model.[60] 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,[61–63] and that its IF is not bounded.[64] Valsecchi *et al.*[65] 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,[66] but many of these proposals rely on residual analysis.[67–70] 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.*[4]

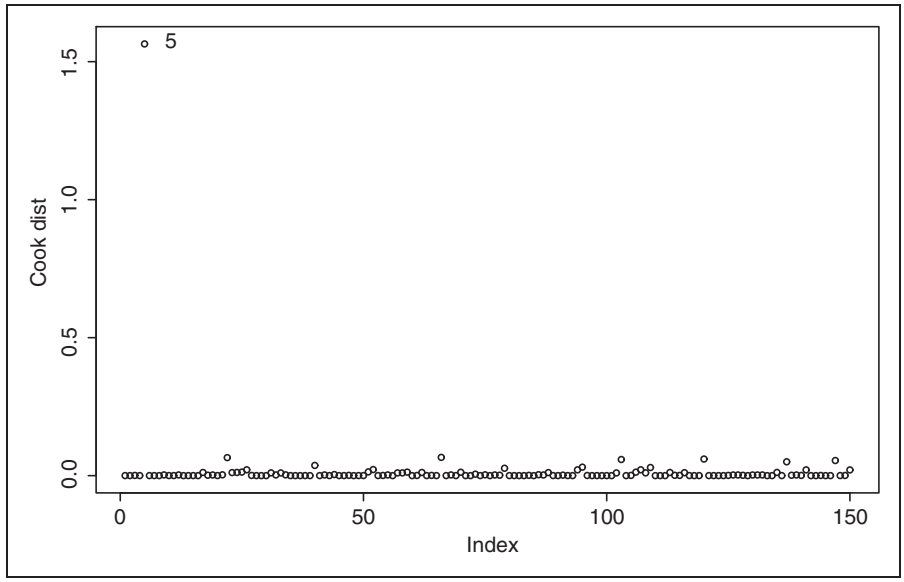

Figure 7. Cook distances for the logistic regression model fit on Food Stamp Data.

Table 4. Summary of robust logistic regression for the Food Stamp Data

	Robust estimate	Standard error	p-Value
β_0	6.49	3.03	0.0321
β_1	−1.83	0.58	0.0018
β_2	0.64	0.55	0.2428
β_3	−1.27	0.52	0.0144

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[60] the effects of covariates on the hazard rate $\lambda(t|x_i)$ for the i-th individual is of the form

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

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

$$S(t|x_i) = \exp\left(-\int_0^t \lambda(u|x_i)\, \mathrm{d}u\right).$$

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

$$L_P(\beta) = \prod_{i=1}^{n} \left(\frac{\exp\left(x_i^T \beta\right)}{\sum_{t_j > t_i} \exp\left(x_j^T \beta\right)} \right)^{\delta_i}. \tag{14}$$

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[71,72] 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;[73] see also Sasieni.[71] 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.[74]

Bednarski,[75] instead, smooths the partial likelihood by introducing weight functions inside the integral equation solved by the MPLE. This idea was refined in Bednarski[76] 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.[77]

Farcomeni and Viviani[78] 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:

$$\begin{cases} \lambda(t|x_i) = \lambda_0(t) \exp(x_i^T \beta) & \text{if } i \in I^* \\ \lambda(t|x_i) = \lambda_i(t) & \text{if } i \notin I^*. \end{cases} \tag{15}$$

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 (15) is the maximiser of

$$L_{\text{TRIM}}(\beta) = \max_{I \in H(\alpha)} \prod_{i \in I} \left(\frac{\exp(x_i^T \beta)}{\sum_{t_j > t_i, j \in I} \exp(x_j^T \beta)} \right)^{\delta_i}. \tag{16}$$

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,[79] Farcomeni and Viviani[78] propose a Metropolis-type algorithm, maximising the

trimmed partial likelihood for fixed α. A non-parametric bootstrap[80] is performed in order to estimate standard errors and CI.

A deep comparison between Bednarski and Sasieni methods can be found in Bednarski and Nowak,[77] while Farcomeni and Viviani[78] 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[70] 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[70] 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.[70]

5.3 An example: prostate cancer data

Data come from Andrews and Herzberg[81] and was used by Nardi and Schemper[70] to illustrate outlier detection in the Cox model.

The prostate cancer study is a randomised trial for prostate cancer comparing a control, given by the use of $\leq$0.2 mg diethylstilbestrol versus a treatment, given by the use of $>$0.2 mg of the same drug. Survival times were recorded for $n = 297$ patients, together with seven binary prognostic factors: an indicator of treatment, performance status (PS, an indicator of limitations of activity), serum haemoglobin level in g/100 mL ($>$12, $\leq$12), weight index ($<$100, $\geq$100), history of cardiovascular disease, tumour size (Small–Large) and a combined index of tumour stage and grade. Censored observations are approximately 33%.

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

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.

Table 5. Summary of Cox model for the prostate cancer data

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
Haemoglobin	−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 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

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*
Haemoglobin	−0.31	−0.43*	−0.29
PS	−0.06	0.36	0.20
Treatment	0.09	0.06	0.049

*Significance at the 5% level.

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 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 of a continuous diagnostic marker. Values of the AUC close to 1 indicate very high diagnostic accuracy, while very low accuracy corresponds to values close to 0.5. The AUC is equal to:[82]

$$A = P(X_1 < X_2), \tag{17}$$

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.

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.*[83] and Pepe.[84] Furthermore, some contributions addressing inference about A have also been provided in semi-parametric and non-parametric settings; see, among others, the recent articles of Adimari and Chiogna[85] and Qin and Jhou.[86]

Assumptions on X_1 and X_2 are of independence and distribution functions $F_1 = F_1(x;\ \theta_1)$ and $F_2 = F_2(x;\ \theta_2)$, respectively. Then, $A = A(\theta) = \int F_1(t;\ \theta_1)\ \mathrm{d}F_2(t;\ \theta_2)$, with $\theta = (\theta_1,\ \theta_2)$. Classical likelihood-based procedures can be badly affected by mild departures from model assumptions, and Greco and Ventura[87] propose a robust inferential procedure to this end. 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 studentised statistic

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

where σ_A^2 is a consistent estimator for the asymptotic variance of $\widehat{A}$.[87]

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 lymphoma (ALCL), which is a rare cancer disease which affects both children and adults. Diseased patients seem to have higher Hsp70 levels than healthy subjects.[88] 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.[89] 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 8. 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[87] 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 $A_{mle} = 0.86$. 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 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

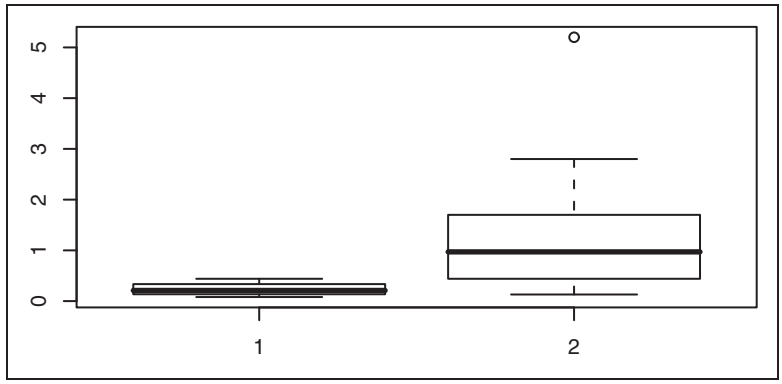

Figure 8. Boxplot of the Hsp70 protein level in controls (1) and cases (2).

Table 7. Point estimates and 95% CI for the AUC in the ALCL lymphoma data

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)

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

7.1 Basic procedures

The simple function `mean` can be used to compute also a trimmed mean, using `mean(x, trim=0.05)`. 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.

7.2 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 BP 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 BP (see Rousseeuw and Leroy,[5] Marazzi[34] and Venables and Ripley,[17] Sec. 6.5).

7.3 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.

7.4 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[78] approach is available from the Web page `http://afarcome.interfree.it/robcox.r`.

Acknowledgements

This work was partially supported by grants from Ministero dell'Universitá e della Ricerca Scientifica e Tecnologica, Italy. The authors acknowledge an anonymous referee for useful comments.

References

1. Huber PJ. *Robust statistics*. New York: Wiley, 1981.
2. Hampel FR, Ronchetti EM, Rousseeuw PJ and Stahel WA. *Robust statistics: the approach based on influence functions*. New York: Wiley, 1986.
3. Maronna RA, Martin RD and Yohai VJ. *Robust statistics: theory and methods*. New York: Wiley, 2006.
4. Heritier S, Cantoni E, Copt S and Victoria-Feser M-P. *Robust methods in biostatistics*. Chichester: Wiley, 2009.
5. Rousseeuw PJ and Leroy AM. *Robust regression and outlier detection*. New York: Wiley-Interscience, 1987.
6. Maronna RA. Principal components and orthogonal regression based on robust scales. *Technometrics* 2005; **47**: 264–273.
7. Hubert M, Rousseeuw PJ and Vanden Branden K. ROBPCA: a new approach to robust principal components analysis. *Technometrics* 2005; **47**: 64–79.
8. Kaufman L and Rousseeuw PJ. *Finding groups in data*. New York: Wiley, 1990.
9. Cuesta-Albertos J, Gordaliza A and Matrán C. Trimmed k-means: an attempt to robustify quantizers. *Ann Stat* 1997; **25**: 553–576.
10. Garcia-Escudero LA and Gordaliza A. Robustness properties of k means and trimmed k means. *J Am Stat Assoc* 1999; **94**: 956–969.
11. Gallegos MT and Ritter G. A robust method for cluster analysis. *Ann Stat* 2005; **33**: 347–380.
12. Farcomeni A. Robust double clustering: a method based on alternating concentration steps. *J Classif* 2009; **26**: 77–101.
13. Hubert M and Van Driessen K. Fast and robust discriminant analysis. *Comput Stat Data Anal* 2004; **45**: 301–320.
14. Copt S and Victoria-Feser MP. High-breakdown inference for mixed linear models. *J Am Stat Assoc* 2006; **101**: 292–300.
15. Huber PJ. Robust estimation of a location parameter. *Ann Math Statist* 1964; **35**: 73–101.
16. Heritier S and Ronchetti E. Robust bounded-influence tests in general parametric models. *J Am Stat Assoc* 1994; **89**: 897–904.
17. Venables WN and Ripley BD. *Modern applied statistics with S*. New York: Springer, 2002.
18. Piron L, Turolla A, Agostini M, Zucconi CS, Ventura L, Tonin P and Dam M. Motor learning principles for rehabilitation: A pilot randomised controlled study in poststroke patients. *Neurorehabil Neural Repair* 2010; **24**: 501–508.
19. Li G. Robust regression. In: Hoagling DC and Tukey JW (eds) *Exploring data tables, trends, and shapes*. New York: Wiley, 1985, pp.281–343.
20. Brazzale AR, Davison AC and Reid N. *Applied asymptotics*. Cambridge: Cambridge University Press, 2007.

21. Cook RD and Weisberg S. *Residuals and influence in regression*. London: Chapman and Hall, 2007.
22. Atkinson AC and Riani M. *Robust diagnostic regression analysis*. New York: Springer-Verlag, 2000.
23. Welsh AH and Ronchetti E. A journey in single steps: robust onestep M-estimation in linear regression. *J Stat Plan Infer* 1985; **103**: 287–310.
24. Bianco AM, Boente G and di Rienzo J. Some results for robust GM-based estimators in heteroscedastic regression models. *J Stat Plan Infer* 2000; **89**: 215–242.
25. Carroll RJ and Ruppert D. Robust estimation in heteroscedastic linear models. *Ann Stat* 1982; **10**: 429–441.
26. Giltinan DM, Carroll RJ and Ruppert D. Some new estimation methods for weighted regression when there are possible outliers. *Technometrics* 1986; **28**: 219–230.
27. Welsh AH and Richardson A. Approaches to the robust estimation of mixed models. In: Maddala G and Rao C (eds) *Handbook of statistics*. Amsterdam: Elsevier, 1997, pp.343–385.
28. Copt S and Heritier S. Robust alternatives to the F-test in mixed linear models based on MM-estimates. *Biometrics* 2007; **63**: 1045–1052.
29. Rousseeuw PJ. Least median of squares regression. *J Am Stat Assoc* 1984; **79**: 871–880.
30. Hubert M, Rousseeuw PJ and Van Aelst S. High-breakdown robust multivariate methods. *Stat Sci* 2008; **23**: 92–119.
31. Carroll RJ and Ruppert D. *Transformation and weighting in regression*. London: Chapman and Hall, 1988.
32. Bellio R. Algorithms for bounded-influence estimation. *Comput Stat Data Anal* 2007; **51**: 2531–2541.
33. Yohai VJ. High breakdown-point and high efficiency robust estimates for regression. *Ann Stat* 1987; **15**: 642–656.
34. Marazzi A. *Algorithms, routines, and S functions for robust statistics*. Pacific Grove: Wadsworth and Brooks/Cole, 1993.
35. Markatou M, Basu A and Lindsay BG. Weighted likelihood equations with bootstrap root search. *J Am Stat Assoc* 1998; **93**: 740–750.
36. Agostinelli C and Markatou M. A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. *Stat Prob Lett* 1998; **37**: 341–350.
37. McCullagh P. Quasi-likelihood and estimating functions. In: Hinkley DV, Reid N and Snell EJ (eds) *Statistical theory and modelling. In Honour of Sir David Cox*. London: Chapman and Hall, 1991, pp.265–286.
38. Adimari G and Ventura L. Quasi-profile loglikelihoods for unbiased estimating functions. *Ann Inst Stat Math* 2002; **54**: 235–244.
39. Bellio R, Greco L and Ventura L. Adjusted quasi-profile likelihoods from estimating functions. *J Stat Plan Infer* 2008; **138**: 3059–3068.
40. Markatou M and Hettmansperger TP. Robust bounded-influence tests in linear models. *J Am Stat Assoc* 1990; **85**: 187–190.
41. Markatou M and Hettmansperger TP. Applications of the asymmetric eigenvalue problem techniques to robust testing. *J Stat Plan Infer* 1990; **31**: 51–65.
42. Hanfelt JJ and Liang KY. Approximate likelihood ratios for general estimating functions. *Biometrika* 1995; **82**: 461–477.
43. Owen AB. *Empirical likelihood*. London: Chapman and Hall, 2001.
44. Adimari G and Ventura L. Quasi-likelihood from M-estimators: a numerical comparison with empirical likelihood. *Stat Meth Appl* 2002; **11**: 175–185.
45. McKean JW, Sheather SJ and Hettmansperger TP. The use and interpretation of residuals based on robust estimation. *J Am Stat Assoc* 1993; **88**: 1254–1263.
46. Rule AD, Larson TS, Bergstralh EJ, Slezak JM, Jacobsen SJ and Cosio FG. Using serum creatinine to estimate glomerular filtration rate: accuracy in good health and in chronic kidney disease. *Ann Int Med* 2004; **141**: 929–938.
47. McCullagh P and Nelder JA. *Generalized linear models*. 2nd ed. London: Chapman and Hall, CRC, 1989.
48. Pregibon D. Resistant fits for some commonly used logistic models with medical applications. *Biometrics* 1982; **38**: 485–498.
49. Copas JB. Binary regression models for contaminated data. *J Roy Stat Soc, Ser B* 1988; **50**: 225–265.
50. Morgenthaler S. Least-absolute-deviations fits for generalised linear models. *Biometrika* 1992; **79**: 747–754.
51. Carroll RJ and Pederson S. On robustness in the logistic regression model. *J Roy Stat Soc, Ser B* 1993; **55**: 693–706.
52. Bianco AM and Yohai VJ. Robust estimation in the logistic regression model. In: Rieder H (ed.) *Robust statistics, data analysis and computer intensive methods*. New York: Springer, 1997, pp.17–34.
53. Markatou M, Basu A and Lindsay B. Weighted likelihood estimating equations: the discrete case with application to logistic regression. *J Stat Plan Infer* 1997; **57**: 215–232.
54. Victoria-Feser M.-P. Robust inference with binary data. *Psychometrica* 2002; **67**: 21–32.
55. Künsch HR, Stefanski LA and Carroll RJ. Conditionally unbiased bounded influence estimation in general regression models, with applications to generalised linear models. *J Am Stat Assoc* 1989; **84**: 460–466.
56. Cantoni E and Ronchetti E. Robust inference for generalised linear models. *J Am Stat Assoc* 2001; **96**: 1022–1030.
57. Wedderburn RWM. Quasi-likelihood function, generalised linear models, and the Gauss-Newton method. *Biometrika* 1974; **61**: 439–447.
58. Heyde CC. *Quasi-likelihood and its application*. New York: Springer-Verlag, 1997.
59. Adimari G and Ventura L. Robust inference for generalized linear models with application to logistic regression. *Stat Prob Lett* 2001; **55**: 413–419.
60. Cox DR. Regression models and life tables (with discussion). *J Roy Stat Soc, Ser B* 1972; **34**: 187–220.
61. Samuels S. *Robustness for survival estimators*. PhD thesis. Department of Biostatistics, University of Washington, 1978.
62. Bednarski T. On sensitivity of Cox's estimator. *Stat Decis* 1989; **7**: 215–228.
63. Minder CE and Bednarski T. A robust method for proportional hazards regression. *Stat Med* 1996; **15**: 1033–1047.
64. Reid N and Crépeau H. Influence function for proportional hazards regression. *Biometrika* 1985; **72**: 1–9.
65. Valsecchi MG, Silvestri D and Sasieni P. Evaluation of long-term survival: use of diagnostics and robust estimators with Cox's proportional hazards models. *Stat Med* 1996; **15**: 2763–2780.
66. Cain KC and Lange NT. Approximate case influence for the proportional hazards regression model with censored data. *Biometrics* 1984; **40**: 493–499.
67. Schoenfeld D. Partial residuals for the proportional hazards regression model. *Biometrika* 1982; **69**: 239–241.
68. Grambsch PM and Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**: 515–526.
69. Therneau TM, Grambsch PM and Fleming T. Martingale based residuals for survival models. *Biometrika* 1990; **77**: 147–160.
70. Nardi A and Schemper M. New residuals for Cox regression and their application to outlier screening. *Biometrics* 1999; **55**: 523–529.

71. Sasieni PD. Maximum weighted partial likelihood estimates for the Cox model. *J Am Stat Assoc* 1993; **88**: 144–152.
72. Sasieni PD. Some new estimators for Cox regression. *Ann Stat* 1993; **21**: 1721–1759.
73. Lin D. Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *J Am Stat Assoc* 1991; **86**: 725–728.
74. Schemper M, Wakounig S and Heinze G. The estimation of average hazard ratios by weighted Cox regression. *Stat Med* 2009; **28**: 2473–2489.
75. Bednarski T. Robust estimation in Cox regression model. *Scand J Stat* 1993; **20**: 213–225.
76. Bednarski T. On a robust modification of Breslow's cumulated hazard estimator. *Comput Stat Data Anal* 2007; **52**: 234–238.
77. Bednarski T and Nowak M. Robustness and efficiency of Sasieni-type estimators in the Cox model. *J Stat Plan Infer* 2003; **115**: 261–272.
78. Farcomeni A and Viviani S. Trimmed Cox regression for robust estimation in survival studies. Technical Report 7, Department of Statistics, Sapienza, University of Rome, 2010.
79. Chakraborty B and Chaudhury P. On an optimization problem in robust statistics. *J Comput Graph Stat* 2008; **17**: 683–702.
80. Davison AC and Hinkley D. *Bootstrap methods and their applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2006.
81. Andrews DF and Herzberg AM. *Data : a collection of problems from many fields for the student and research worker*. New York: Springer-Verlag, 1985.
82. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psy- col* 1975; **12**: 387–415.
83. Kotz S, Lumelskii Y and Pensky M. *The stress-strength model and its generalizations: theory and applications*. Singapore: World Scientific, 2003.
84. Pepe MS. *The statistical evaluation of medical tests for classification and prediction*. Oxford: Oxford University Press, 2003.
85. Adimari G and Chiogna M. Partially parametric interval estimation of Pr(Y>X). *Comput Stat Data Analysis* 2001; **51**: 1875–1891.
86. Qin GS and Zhou XH. Empirical likelihood inference for the area under the roc curve. *Biometrics* 2006; **62**: 613–622.
87. Greco L, Ventura L. Robust inference for the stress-strength reliability. Statistical Papers, 2010.
88. Mayer MP and Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* 2005; **62**: 670–684.
89. Cortese G and Ventura L. *Accurate likelihood on the area under the ROC curve for small samples*. Technical Report 17, Department of Statistics, University of Padova, 2009.