How to Marry Robustness and Applied Statistics

Andrea Cerioli, Anthony C. Atkinson and Marco Riani

Abstract

A striking feature of most applied statistical analyses is the use of methods that are well known to be sensitive to outliers or to other departures from the postulated model. Since data contamination is often the rule, rather than the exception, we investigate the reasons for this contradictory (and perhaps unintended) choice. We also provide empirical evidence, in a real-world regression problem concerning international trade, of the advantages of a new approach to data analysis based on monitoring. Our approach enhances the applicability of robust techniques and the interpretation of their results, thus yielding a positive step towards a reconciliation between robustness and applied statistics.

1 Introduction

An early use of the term robustness is due to [5] in a study of the effect of non-normality on tests of equality of variances. He commented that means are robust to departures from normality, but that estimates of variances are not. The matter

A. Cerioli (✉) · M. Riani
Dipartimento di Economia, Università di Parma, Parma, Italy
e-mail: andrea.cerioli@unipr.it

M. Riani
e-mail: mriani@unipr.it

A.C. Atkinson
Department of Statistics, School of Economics and Political Science, London, UK
e-mail: a.c.atkinson@lse.ac.uk

T. Di Battista et al. (eds.), *Topics on Methodological and Applied Statistical Inference*, Studies in Theoretical and Applied Statistics,
DOI 10.1007/978-3-319-44093-4_6

is clearly important, since data frequently depart from the assumptions behind the models of mathematical statistics used to derive tests and other statistical procedures. Twenty years after Box, the outlines of the modern theory of robust statistics were becoming clearly established as the development of procedures that behaved well under small departures from standard assumptions, typically those of normality. This is a much narrower study than that implied by Box. One purpose of our paper is to show how the range of application of robust methods can be extended through the use of 'monitoring', exemplified in Sect. 3, where we study aspects of fitted models under a series of assumptions about the level of contamination in the data. An important byproduct is a simplification of the numerous choices required in the application of robust methods. We conclude with a discussion of problems that have mostly not been the subject of robust analysis. One example, treated in Sect. 4.1, is the identification of data that not only include outliers but in which groups of observations come from different models. But we start with a brief history of the development of robust methods.

Reference [38] gives a short history of robustness. The earliest book-length reference is [1] (the Princeton Robustness Study), at which time it was expected that all statistical analyses would, by default, be robust. Now, a further forty years on, there are at least 6 books about robust statistics with over 1,000 citations in Google Scholar. At the time of writing, the most highly cited is [23] (and its second edition [25]), the others, in citation order, are [1,18,21,28,36]. Unfortunately, this activity seems largely to be statisticians talking to other statisticians. Recent references with more applied emphasis, especially to health sciences, include [15,20] and the forthcoming book by [14]. However, the impact of robust methodologies in substantive domains still remains minor.

Although the term "robust" was first popularized by Box, the idea of considering the distribution of statistics under departures from the assumption of normality goes back at least to E.S. Pearson's review [29] of the second edition of Fisher's *Statistical Methods for Research Workers*, an interest that was to stay with Pearson until the end of his scientific life. The current understanding of robust was much more the creation of Tukey, starting with [39], and of [22]. Stigler writes:

"... by 1972 a number of the early workers in robust statistics expected that from the 1970s to 2000 we would see the same development with robust methods—extensions to linear models, time series, and multivariate models, and widespread adoption to the point where every statistical package would take the robust method as the default.... This was, and I [Stigler] will call it, a Grand Plan. But that plainly is not what has occurred".

Stigler then presents a lively and warm discussion of the early history of robust statistics. One reason for the lack of opening to the scientific world may be that robust statistics, as often understood and practised, has led to a new mathematical statistics, more complicated than the old, in which ever more refined solutions are presented to a few well-defined problems. We describe some of these complications in Sect. 2. From the standpoint of a user of statistical methods, the result of a robust analysis is to provide an alternative, for example for regression, to least squares. There are therefore two summaries of the data, rather than one. That this is not an especially appealing development is evidenced by the failure of major commercial statistical packages to implement robust methods of data analysis except as special procedures within a well-segregated collection of routines. We appreciate that there are several robust libraries available in R, but would argue that, again, this package tends to be statisticians talking to statisticians.

Stigler suggests that the first signs of trouble with the Grand Plan were already evident in 1972 at the time of the publication of the Princeton Robustness Study. To quote him again:

"From the full set of 10,465 estimates of a location parameter they had considered, they reported in detail on the accuracy of 68 estimates that had received extensive study, focusing upon small samples and an inventively wide selection of 32 distributions, nearly all of which were symmetric scale mixtures of normal distributions".

This is far from the Grand Plan and, indeed, none of the authors of the conclusions in Chap. 7 of the Study made any grand claims for their work. Unlike the psalmist, they all display a compulsive refusal to lift their eyes to the hills, even for a moment; no Grand Plan is needed. But the year's work in Princeton by many intellectually impressive statisticians did not move far in solving the typical problems of data analysis mentioned in our first sentences. Indeed, your second author remembers the mounting despair with which a reading party organized by David Cox at Imperial College worked through the Study. We were quickly mired in the details of trying to remember what was, for example, an 'iteratively C-skipped trimean'. In Cox's recent book on applied statistics [13] the index contains just one reference to robustness. The relevant page carefully discusses the identification and treatment of outliers, stressing the comparative difficulty of the identification of multiple outliers and the importance of considering physical interpretation for any outliers found; points partially illustrated in the analysis of our example in Sect. 3. Likewise, the main reference to robustness in Huber's recent book on data analysis [24] downplays formal methods of robustness. In Sect. 5.3, 'Mathematical statistics and approximate models' Huber writes about the work of Fisher that, after Fisher "the robustness paradigm – explicitly permitting small deviations from the idealized model when optimizing – carried [the argument] only a few steps further". We hope to show that these works underestimate the contribution to intelligent data analysis that can be made by proper monitoring of the robust methods developed over 50 years since the study.

The most extreme forms of robustness usually considered are a very robust fit, asymptotically resistant to 50 % of aberrant observations, and maximum likelihood, including least squares, which have zero breakdown point. It is common [36,37] to suggest comparison of the residuals or Mahalanobis distances from such fits. In the approach illustrated in Sect. 3 we extend this idea, monitoring such quantities as residuals or distances, parameter estimates, test statistics and other quantities of interest as the robustness of the fit decreases. We thus obtain information on important changes in conclusions that come from differing assumptions about the degree of contamination in the data.

One consequence of our monitoring of robust procedures is that, by considering a variety of procedures for robust fitting, we are able to determine which, amongst the many parameters of the algorithms, are those that are critical, distinguishing them from those that are only of secondary importance. The final goal is to provide insightful data analyses by following well-specified procedures that can be straightforwardly applied by non-specialists in robust statistics.

Our paper is structured as follows: in Sect. 2 we discuss the choice of an appropriate form of robust method, with an emphasis on regression, difficulties in numerical procedures and the interpretation of the results of a robust analysis. An important statistical drawback to downweighting methods, as opposed to trimming, is the breaking of the connection between each observational unit and quantities derived from the analysis, such as parameter estimates.

An example of monitoring is in Sect. 3 where we compare two methods of robust regression. One, S estimation, reveals that robust and non-robust fits to the data are very different; the other method, MM estimation, fails to do so, a finding in line with the conclusions of [33], who use monitoring to compare many different forms of robust regression.

As the quotation above from [24] indicates, standard robust methods have typically been developed under the assumption that there is a single model from which there are small departures, such as a slightly non-normal distribution of errors, perhaps together with gross outliers. This is only a slight part of the broad range of possible departures the data analyst may face. We indicate many such problems in Sect. 4, the theme of which is "robustness against what"? One important form of departure arises when the data are a mixture of observations from more than one model. For multivariate normal populations, this leads to problems of clustering. In Sect. 4.1 we continue the analysis of the regression data from Sect. 3, showing that they come from two different regression models. There are also a number of outliers. An important feature of robust clustering is that it is not necessary to cluster all observations. Our random start method based on the Forward Search does not require prior specification of the amount of trimming required, a feature it shares with the method of monitoring of Sect. 3. The subsequent section of the paper discusses some related issues that may contribute to discouraging the use of robust techniques, such as the difficulty in obtaining a reliable estimate of the number of outliers and the lack of knowledge about the empirical behaviour of the methods when the errors are very non-normal.

2 Which Method and How to Tune It?

A major disincentive to the routine use of standard robust methods is the number of decisions that have to be made before the analysis of the data begins. We now describe some of these.

1. The efficient application of robust methods depends on the proportion of outliers expected in the particular set of data being analysed. These determine the desired efficiency or, equivalently, breakdown point. Clearly, a very robust analysis can

always be used, but this results in an unnecessarily low efficiency for data that are virtually outlier free.

2. The next choice is the nature of robust estimator that is required. For regression [33] identify three classes of estimators:

    a. Hard (0,1) trimming such as Least Trimmed Squares - LTS: [17,35] or Least Median of Squares - LMS: [35] in which the amount of trimming is determined by the choice of the trimming parameter.
    b. Adaptive Hard Trimming. In the Forward Search (FS), the observations are again hard trimmed, but the amount of trimming is determined by the data, being found adaptively by the search. See [2,32] for regression and [4] for a general survey of the FS, with discussion.
    c. Soft trimming (downweighting). M estimation and derived methods, including weighted likelihood. The intention is that observations near the centre of the distribution essentially retain their value, but a suitable weight function ensures that increasingly remote observations have an effect on fitting that decreases with distance from the centre [25,27,28].

3. Within the soft trimming family, both the weight function, often called $\rho(\cdot)$, and the one or two parameters determining efficiency have to be chosen. Reference [33] use monitoring to compare three methods: S, MM and τ for four ρ functions: Tukey's bisquare, optimal, Hyperbolic and Hampel.

The calculations for robust estimation are also much more difficult than those of least squares. The functions to be maximized when using robust estimators are typically complicated, with many local maxima. In consequence, approximate methods are used. The standard approach uses randomly sampled subsets of p observations (elemental sets). We now list some of the choices that have to be made to provide a viable algorithm.

1. The number of subsamples to extract, to each of which the model is fitted exactly. These fitted values are used to evaluate the function to be maximized.
2. The maximum number of refining iterations (concentration steps), if any, within each subsample.
3. The tolerance for the convergence of the estimate of β in the refining steps.
4. The number of best subsets resulting from the refining steps to be brought to convergence.
5. The number of refining iterations for each best subset being brought to convergence.
6. The tolerance for the estimate of β in the refining steps for each subset being brought to convergence.
7. The tolerance for the estimate of scale in the best subsets.

In calculations for the example in Sect. 3 we follow the recommendations of the FSDA toolbox. Reference [19] show that inappropriate choices of some of these tuning constants may lead to inconsistency of the resulting algorithms.

Perhaps even more important than these technical matters, are the statistical problems related to, in particular, the downweighting of observations.

1. There is a loss of simplicity in the tests related to parameter estimates. In their Sect. 7.6, [25] point out that there may be alternative and equally plausible robust variants of the asymptotic standard errors of estimated regression coefficients. Reference [33] describe, and later exemplify, two such robust variants of the usual t-test, which sometimes differ in the conclusions they lead to. There is no guidance as to which is to be preferred in such circumstances.
2. Through the use of downweighting, the analyst loses the connection between each unit and the parameter estimates and other statistically important quantities. We note that this connection is maintained in the FS and other hard trimming methods.

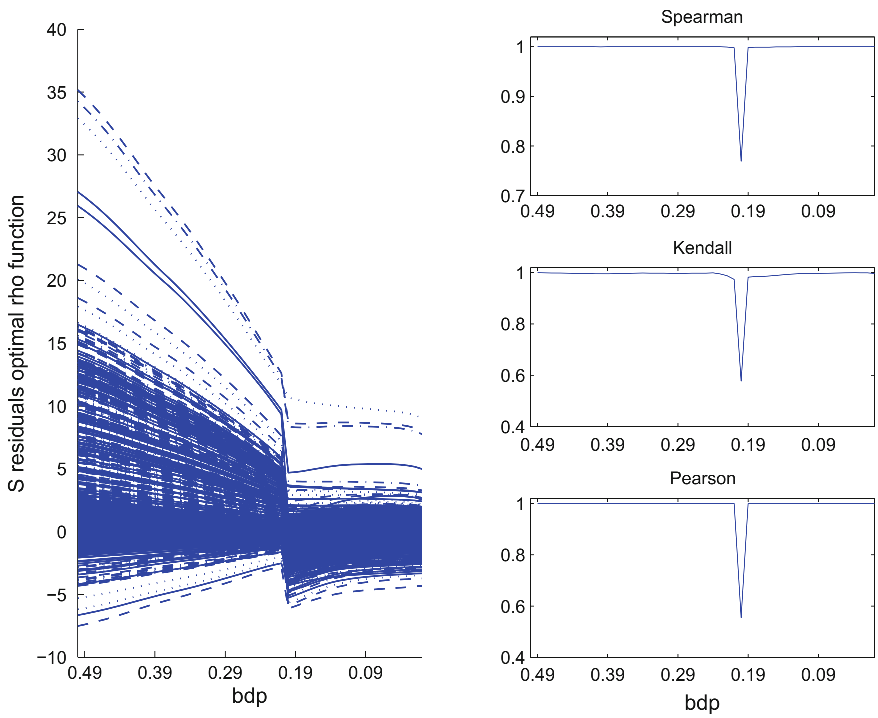

Fig. 1 Vegetable products data. S estimation, optimal ρ function. *Left-hand panel*, plot of scaled residuals. *Right-hand panel*, three measures of the correlations of adjacent residuals. The abrupt switch virtually to LS at 0.20 is evident in both panels

3 An Example of Monitoring

We illustrate the use of monitoring in the context of international trade, which is an important field of application for the EU economy. For instance, [8] describe the importance of careful statistical analysis of international trade data and some of the challenges emerging in such an exercise. The dataset that we consider contains the value and weight of $n = 1{,}558$ import transactions of vegetable products, such as oils and seeds, to one specific EU Member State from a non-EU country. To illustrate the usefulness of monitoring in understanding the properties of various robust estimators, we compare S and MM estimation. Typically we require 50 robust regression fits per analysis; a computational burden only made possible by the efficiency of the FSDA robust library [31] and by the recent technical advances of [34].

In monitoring S estimators we vary the bdp from 0.5 to 0.01. For MM estimates it is more convenient to monitor changes as the efficiency goes from 0.5 to 0.99. In both cases we look at plots of all n residuals as a function of efficiency or bdp. A useful diagnostic, summarizing the plot of residuals, is to plot correlation of the ranks between the residuals at adjacent monitoring values. We consider three standard measures of correlation:

1. Spearman. The correlations between the ranks of the two sets of observations.
2. Kendall. Concordance of the pairs of ranks.
3. Pearson. Product-moment correlation coefficient.

If there is a clear division of the solutions into a robust fit and a non-robust one, with a sharp break between them, this is clearly shown by the correlation plot. For more complicated examples the point of transition is not so clearly visible. But the structure of the residual plot is well summarized by looking at correlations.

Figure 1 shows the plot of residuals for S estimation. There is a clear break in the plot between bdp 0.21 and 0.20, as the robust fit changes to least squares. For the LS fit there seems to be an almost symmetrical distribution of residuals, with around half a dozen large positive outliers. The robust fit, for higher values of the bdp, exhibits a highly skewed structure for the residuals. The constancy of the ranking of the residuals over the two regimes is clearly shown in the right-hand panel of the plot; all three correlations are virtually one, except for the break point between the bdp of 0.21 and 0.20.

This figure is very different from that for MM estimation in Fig. 2. Here the pattern of residuals is constant for all efficiencies in the range studied and similar to that for the robust part of the S residuals in Fig. 1. The correlation plots show no change in the pattern.

These results show an appreciable difference between S estimation and MM, which is tuned to have a high efficiency for the parameters of the linear model. We now explore the parameter estimates of the linear model and their relationship with the data.

In these data there is a single explanatory variable. Figure 3 shows how the estimate of the slope changes with the bdp for S estimation and the efficiency for MM

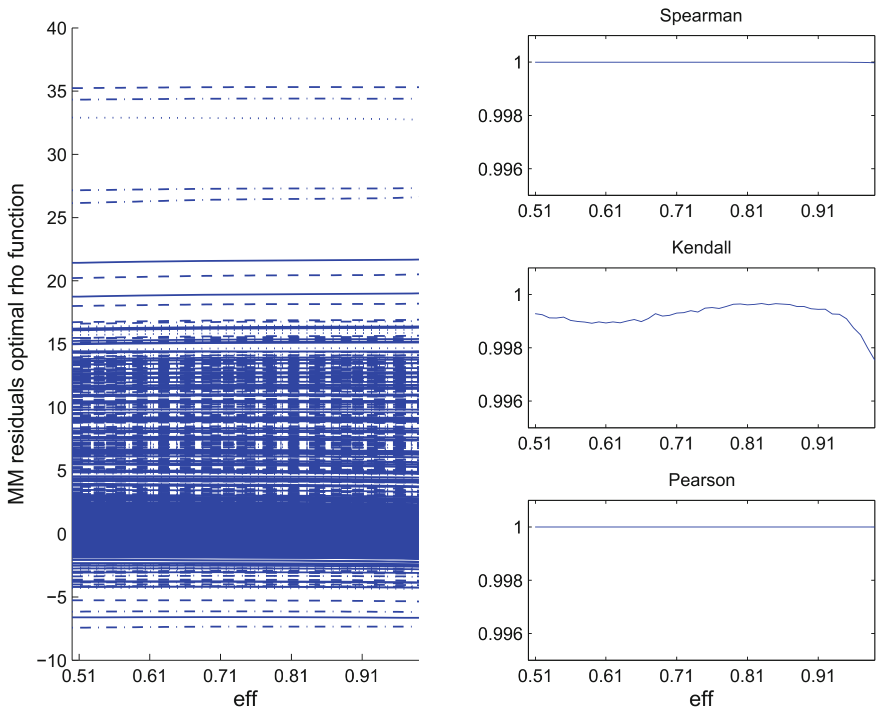

Fig. 2 Vegetable products data. MM estimation, optimal ρ function. *Left-hand panel*, plot of scaled residuals. *Right-hand panel*, three measures of the correlations of adjacent residuals. The skewed distribution of residuals remains constant over the considered range of efficiency. There is no change in the values of the correlation coefficients in the *right-hand panel* (note the vertical scale)

estimation. For S estimation the slope remains virtually constant, decreasing from 3.42 to 3.32, until, with a bdp of 0.20, it jumps up to 5.95. Thereafter it again decreases slightly, with a minimum value of 5.18. On the contrary, for MM estimation the slope decreases slowly from 3.40 to 3.32; the jump in values is missing.

The behaviour of the S slope that is revealed by monitoring is what might be expected if there is a main population following a regression line and a cluster of outliers at a position of high leverage. The right-hand panel of Fig. 3 shows five fitted lines for S estimation. Those for high bdp down to 0.21 follow the lower line of data. The fit calculated with bdp $= 0.20$ lies close to the upper line, for which there are more observations than for the lower one. As the bdp further decreases the lines move slightly towards lying between the two main lines, being attracted upwards by the presence of a few large disconnected outliers, some with appreciable leverage. The plot for MM lies throughout close to the lower line.

The conclusion from this analysis is that monitoring using S estimation alerts us to the presence of a structure in the data that would not be so trenchantly revealed by looking at the output from a single fit. Monitoring MM estimation, on the other

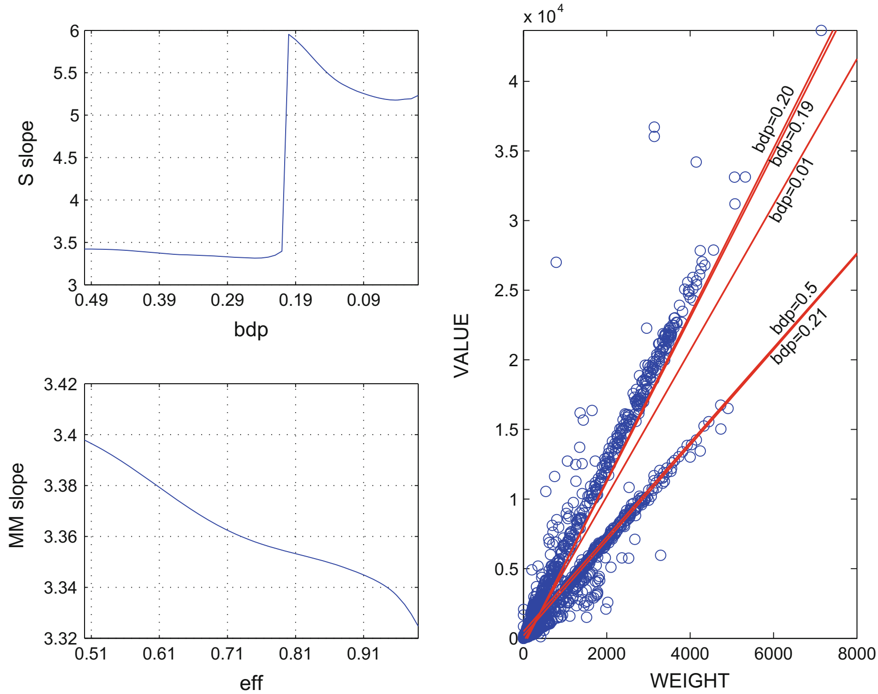

Fig. 3 Vegetable products data. *Left-hand panels*, the estimated slope parameters for *S* estimation (*upper panel*) and *MM* estimation (*lower panel*). *Right-hand panel*, the data and fitted lines using *S* estimation with five different values of *bdp*

hand, does not indicate that there is an important departure from the single model assumed to hold for the majority of the data. Perusal of Fig. 2 might, on the contrary, suggest that a transformation of the data is needed to achieve a symmetrical error distribution.

The results here from the comparison of S and MM estimation are in line with those of the extended study of this kind of monitoring by [33] who conclude that highly tuned methods like MM and τ estimation often reveal less about the structure of the data than does S estimation. Of the four ρ functions they compare, they show that Tukey's bisquare and the closely related optimal function provide the most informative monitoring. The hyperbolic ρ function, for some sets of data, is subject to numerical problems. Here we have used the optimal function.

We return to these data in the next section. Before we do so, we note that it might be expected that fitted lines for value against weight should go through zero. We did repeat our analysis setting the regression intercept to zero, but found that the conclusions were unaffected. Although, in some trading activities, there is a non-zero intercept, being the cost of setting up an order, such an effect is more common in domestic mail orders than in the kind of data we are analysing here.

4 Robustness Against What?

Standard robust methods were developed for fitting a single model. In this section we first describe a robust method for determining whether the data are a mixture from more than one model, although there is the restriction that the models are all of the same class. In the subsequent section we briefly discuss the more general, and far broader, problem of robustness when the class of model, or models, also needs to be identified.

4.1 Several Models: Clustering

The analysis of the trade data in Sect. 3 with monitoring shows that the robust S fit and least squares differ. However there is no clear indication of what is causing the difference. Of course, with a single explanatory variable, a simple scatterplot indicates the structure. But, in general, there may be several explanatory variables or so much data that perusal of individual scatterplots for all types of transaction is impossible. We use the FS to provide a robust analysis of data when there are several sources for the data. We need a robust method as we need to avoid the deleterious effect of the outliers, the presence of which is evident in the figure.

The forward search achieves robustness by fitting the model to subsets of the data of increasing size, where the subsets are sequentially chosen to contain observations as close as possible to the fitted model. The introduction of outliers into the subset is diagnostically revealed by plots of residuals against subset size as well as formally by statistically tuned tests using the minimum deletion residual among observations not in the subset. The method for a single population starts from a robustly chosen subset of m_0 observations. However, if the data are a mixture of observations generated from more than one model, the robustly chosen initial subset $S^*(m_0)$ may lead to a search in which observations from several models enter the subset haphazardly in such a way that the various models are not revealed. Searches from more than one starting point are necessary to reveal the more complicated structure of a mixture.

For finding clusters in multivariate data, [3] suggest running several hundred searches from randomly chosen initial subsets m_0. At the beginning of the search with regression models, a random start produces some very large residuals. But, because the search can drop units from the subset as well as adding them, some searches are attracted to specific regression lines. As the searches progress, the various random start trajectories converge, with subsets containing the same units. Once trajectories have converged, they cannot diverge again. As we see in Fig. 4, which is typical of those for many data structures, the search is rapidly reduced to relatively few trajectories, some of which show marked peaks. It is these that provide information on the number and membership of the clusters.

The two peaks in Fig. 4 indicate the two linear structures that are apparent in Fig. 3. The final peak in the plot results from the outliers, which are also evident in Fig. 3. The next step in the analysis of the data is to 'interrogate' the peaks, taking many of the units in the subset just before the peaks as large initial subsets for forward searches to confirm cluster membership. The availability of automatic procedures for deciding cluster membership is an advance over many robust clustering procedures which require prior information on the number of clusters and on the proportion of the data to be trimmed, and so suffer from one of the main disadvantages of robust methods listed in Sect. 2.

A final word is in order about the interpretation of the forward plot of deletion residuals in Fig. 4. In all there are 1,558 observations. However, the two peaks come at $m = 1{,}174$ and 1,310, which total much more than all the observations. There are, however, an appreciable number of observations at low values of x; due to variability in the data, these could belong to either line. Straightforward clustering would be unable to decide to which line such observations should be allocated.

4.2 Which Model for the 'Good' Data and How Many Outliers?

The development of high-breakdown techniques, like S and MM estimation, has been the mainstream of theoretical work on robust statistics for at least 25 years. These methods are expected to work well in a contamination framework where the data generating distribution, say $G(y)$, is such that

$$G(y) = (1-\gamma)G_0(y) + \gamma G_1(y). \tag{1}$$

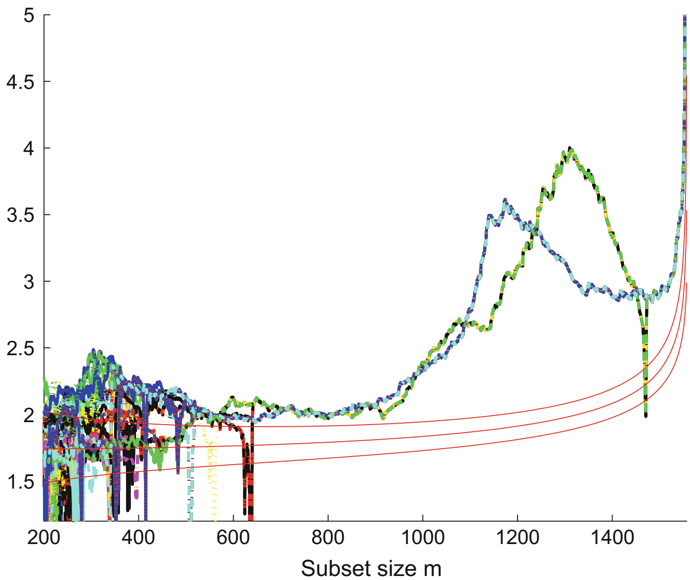

Fig. 4 Vegetable products data: forward plots of minimum deletion residuals from 200 random starts with pointwise 1 and 99 % limits. There appear to be two distinct groups (regression lines)

In model (1), $G_0(y)$ and $G_1(y)$ denote the distribution functions of the 'good' and of the contaminated part of the data, respectively, and $\gamma < 0.5$ is the unknown contamination rate.

We speculate that another reason for the limited appeal of robust methods in practical applications is the need to specify $G_0(y)$. Furthermore, very little is known about both the theoretical and empirical behaviour of the techniques when $G_0(y)$ is not normal. To motivate our claim, we observe that all high-breakdown estimators require computation of a normalizing constant which ensures consistency when $\gamma = 0$. In the case of hard trimming, this constant is a scaling factor for the estimate of dispersion and, in the case of soft trimming, a threshold above which observations are given zero weight. As far as we know, explicit and computable formulae for the normalizing constant exist only if $G_0(y)$ is the normal distribution and, indeed, relevant real-world applications have been confined to this model.

Reference [10] propose a method for testing the hypothesis that $G_0(y)$ in (1) is normal. The good power properties of their test seem to suggest that the empirical behaviour of high-breakdown techniques may be considerably different under non-normal models, especially when $G_0(y)$ is skewed. Furthermore, they show the potentially deleterious consequences of a naive approach to robustness which is often implemented in practice, when standard methods are applied to the observations that remain after outlier removal.

Even when $G_0(y)$ is the normal distribution, many high-breakdown procedures show poor finite sample properties for estimation of the contamination rate γ. The tendency to produce a plethora of spurious outliers has been shown in many studies, starting from [12] and including [9]. We argue that this tendency has also been a serious constraint on the dissemination of robust methods among practitioners. As a consequence, we strongly advocate the use of robust techniques that are able to provide effective control on the number of false discoveries, while keeping good detection properties. References [6,7] propose modified high-breakdown procedures that can achieve this goal, while [30,33] and this paper point towards a flexible monitoring approach.

5 Conclusion

We argue that there is compelling need for a reconciliation between robustness and applied statistics. In this paper we have investigated some of the reasons that we see as major disincentives to the routine use of standard robust methods. We have also provided empirical evidence, in a regression setting and in a real-world problem concerning international trade, of the advantages of a new approach to data analysis based on monitoring.

We conclude by noting that our monitoring approach deserves further theoretical investigation. A pioneering contribution in this direction, although in a somewhat simplified setting, is the study of the asymptotic properties of the radius process of [16]. Results for the forward search are provided by [11,26], while the properties of the trajectories of the residuals computed from other high-breakdown estimators, like those given in Figs. 1 and 2, are still unexplored. Nevertheless, we trust that our work will provide a positive contribution towards the desired reconciliation.

Acknowledgments We thank the Scientific Program Committee of the 47th Scientific Meeting of the Italian Statistical Society for inviting us to present this work. We are also grateful to Dr. Domenico Perrotta of the European Commission Joint Research Centre at Ispra for providing the data on trade in vegetable products. Our work on this paper was partly supported by the project MIUR PRIN "*MISURA – Multivariate models for risk assessment*".

References

1. Andrews, D.F., Bickel, P.J., Hampel, F.R., Tukey, W.J., Huber, P.J.: Robust Estimates of Location: Survey and Advances. Princeton University Press, Princeton (1972)
2. Atkinson, A.C., Riani, M.: Robust Diagnostic Regression Analysis. Springer, New York (2000)
3. Atkinson, A.C., Riani, M., Cerioli, A.: Monitoring random start forward searches for multivariate data. In: Brito, P. (ed.) COMPSTAT, pp. 447–458. Physica-Verlag, Heidelberg (2008)
4. Atkinson, A.C., Riani, M., Cerioli, A.: The forward search: theory and data analysis (with discussion). J. Korean Stat. Soc. **39**, 117–134 (2010)
5. Box, G.E.P.: Non-normality and tests on variances. Biometrika **40**, 318–335 (1953)
6. Cerioli, A.: Multivariate outlier detection with high-breakdown estimators. J. Am. Stat. Assoc. **105**, 147–156 (2010)
7. Cerioli, A., Farcomeni, A.: Error rates for multivariate outlier detection. Comput. Stat. Data Anal. **55**, 544–553 (2011)
8. Cerioli, A., Perrotta, D.: Robust clustering around regression lines with high density regions. Adv. Data Anal. Classif. **8**, 5–26 (2014)
9. Cerioli, A., Riani, M., Atkinson, A.C.: Controlling the size of multivariate outlier tests with the MCD estimator of scatter. Stat. Comput. **19**, 341–353 (2009)
10. Cerioli, A., Farcomeni, A., Riani, M.: Robust distances for outlier-free goodness-of-fit testing. Comput. Stat. Data Anal. **65**, 29–45 (2013)
11. Cerioli, A., Farcomeni, A., Riani, M.: Strong consistency and robusteness of the forward search estimator of multivariate location and scatter. J. Multivar. Anal. **126**, 167–183 (2014)
12. Cook, R.D., Hawkins, D.M.: Comment on Rousseeuw and van Zomeren. J. Am. Stat. Assoc. **85**, 640–644 (1990)
13. Cox, D.R., Donnelly, C.A.: Principles of Applied Statistics. Cambridge University Press, Cambridge (2011)
14. Farcomeni, A., Greco, L.: Robust Methods for Data Reduction. Chapman and Hall/CRC, Boca Raton (2015)
15. Farcomeni, A., Ventura, L.: An overview of robust methods in medical research. Stat. Methods Med. Res. **21**, 111–133 (2012)
16. García-Escudero, L.A., Gordaliza, A.: Generalized radius processes for elliptically contoured distributions. J. Am. Stat. Assoc. **100**, 1036–1045 (2005)
17. Hampel, F.R.: Beyond location parameters: robust concepts and methods. Bull. Int. Stat. Inst. **46**, 375–382 (1975)
18. Hampel, F., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A.: Robust Statistics. Wiley, New York (1986)
19. Hawkins, D.M., Olive, D.J.: Inconsistency of resampling algorithms for high-breakdown regression estimators and a new algorithm (with discussion). J. Am. Stat. Assoc. **97**, 136–159 (2002)

20. Heritier, S., Cantoni, E., Copt, S., Victoria-Feser, M.P.: Robust Methods in Biostatistics. Wiley, Chichester (2009)
21. Hoaglin, D.C., Mosteller, F., Tukey, J.W.: Understanding Robust and Exploratory Data Analysis. Wiley, New York (1983)
22. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Stat. **35**, 73–101 (1964)
23. Huber, P.J.: Robust Statistics. Wiley, New York (1981)
24. Huber, P.J.: Data Analysis: What Can be Learned from the Past 50 Years. Wiley, New York (2011)
25. Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn. Wiley, New York (2009)
26. Johansen, S., Nielsen, B.: Analysis of the forward search using some new results for martingales and empirical processes. Bernoulli **21** (2015, in press)
27. Markatou, M., Basu, A., Lindsay, B.G.: Weighted likelihood estimating equations with a bootstrap root search. J. Am. Stat. Assoc. **93**, 740–750 (1998)
28. Maronna, R.A., Martin, R.D., Yohai, V.J.: Robust Statistics: Theory and Methods. Wiley, Chichester (2006)
29. Pearson, E.S.: Statistics in biological research. Nature **123**, 866–867 (1929)
30. Riani, M., Atkinson, A.C., Cerioli, A.: Finding an unknown number of multivariate outliers. J. R. Stat. Soc. Ser. B **71**, 447–466 (2009)
31. Riani, M., Perrotta, D., Torti, F.: FSDA: a MATLAB toolbox for robust analysis and interactive data exploration. Chemom. Intell. Lab. Syst. **116**, 17–32 (2012)
32. Riani, M., Atkinson, A.C., Perrotta, D.: A parametric framework for the comparison of methods of very robust regression. Stat. Sci. **29**, 128–143 (2014)
33. Riani, M., Cerioli, A., Atkinson, A.C., Perrotta, D.: Monitoring robust regression. Electron. J. Stat. **8**, 646–677 (2014)
34. Riani, M., Cerioli, A., Torti, F.: On consistency factors and efficiency of robust S-estimators. TEST **23**, 356–387 (2014)
35. Rousseeuw, P.J.: Least median of squares regression. J. Am. Stat. Assoc. **79**, 871–880 (1984)
36. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley, New York (1987)
37. Rousseeuw, P.J., van Zomeren, B.C.: Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc. **85**, 633–639 (1990)
38. Stigler, S.M.: The changing history of robustness. Am. Stat. **64**, 277–281 (2010)
39. Tukey, J.W.: A survey of sampling from contaminated distributions. In: Olkin, I., et al. (eds.) Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, pp. 448–485. Stanford University Press, Palo Alto (1960)