Robust Transformations for Multiple Regression via Additivity and Variance Stabilization

Marco Riani, Anthony C. Atkinson & Aldo Corbellini

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

Robust Transformations for Multiple Regression via Additivity and Variance Stabilization

Marco Riani [a], Anthony C. Atkinson [b], and Aldo Corbellini [a]

aDipartimento di Scienze, Economiche e Aziendale and Interdepartmental Centre for Robust Statistics, Università di Parma, Parma, Italy; bThe London School of Economics, London, UK

ABSTRACT

Outliers can have a major effect on the estimated transformation of the response in linear regression models, as they can on the estimates of the coefficients of the fitted model. The effect is more extreme in the Generalized Additive Models (GAMs) that are the subject of this article, as the forms of terms in the model can also be affected. We develop, describe and illustrate robust methods for the nonparametric transformation of the response and estimation of the terms in the model. Numerical integration is used to calculate the estimated variance stabilizing transformation. Robust regression provides outlier free input to the polynomial smoothers used in the calculation of the response transformation and in the backfitting algorithm for estimation of the functions of the GAM. Our starting point was the AVAS (Additivity and VAriance Stabilization) algorithm of Tibshirani. Even if robustness is not required, we have made four further general optional improvements to AVAS which greatly improve the performance of Tibshirani's original Fortran program.

We provide a publicly available and fully documented interactive program for our procedure which is a robust form of Tibshirani's AVAS that allows many forms of robust regression. We illustrate the efficacy of our procedure through data analyses. A refinement of the backfitting algorithm has interesting implications for robust model selection. Supplementary materials for this article are available online.

1. Introduction

The nonlinear transformation of response variables is a common practice in regression problems. Two customary goals are the stabilization of error variance and the approximate normalization of the error distribution (Box and Cox 1964). The more comprehensive goals of our article are to find, via smoothing and robustness, those transformations that produce the best-fitting additive model. Hopefully, such knowledge may lead to parsimonious parametric models capable of a physical interpretation.

Generalized Additive Models (GAMs) are a widely employed extension of linear regression models which we use to replace the explanatory variables by nonparametrically estimated functions of the variables (Hastie and Tibshirani 1990). Tibshirani (1988) extended the GAM to include response transformation using an empirical version of the variance stabilizing transformation. His algorithm AVAS (additivity and variance stabilization) provides nonparametric alternatives to the Box-Cox parametric transformation of the response and the Box-Tidwell family of power transformations of explanatory variables (Box and Tidwell 1962). Unfortunately, like these parametric methods, his procedure is not robust; the transformations of both the explanatory variables and the response can be highly distorted by outlying observations (Atkinson, Riani, and Corbellini 2021). The same problem holds for the related ACE algorithm of Breiman and Friedman (1985) and is one topic in the discussion of that

article (Fowlkes and Kettenring 1985; Buja and Kass 1985). It is our purpose to provide a robust version of AVAS, robust both for transformations of the response and the explanatory variables. For obvious reasons we call our procedure RAVAS.

Both ACE and AVAS are appreciably cited and applied and both are available in the R package `acepack` (Spector et al. 2016). Although Breiman and Friedman (1985) present ACE as a method for transformations in multiple regression and correlation, it has some anomalous properties for estimation of response transformations, as noted by Tibshirani (1988). A discussion of the relationship of ACE and AVAS and of both to the Box-Cox transformation is in Hastie and Tibshirani (1990, Cap.7). Buja and Kass (1985) suggest a method for the robustification of ACE, but comment that "it is far easier to make this glib remark than to formulate the problem in such a way that progress [can] be made while retaining such advantages of the current ACE algorithm as the low computational cost." The same remark applies to AVAS. What is surprising is the paucity of references for improving the procedure. Foi (2008) comments that AVAS is non-asymptotic. Both he and Wang, Lyu, and Yu (2021) use function minimization for numerical variance stabilization in an image-enhancement problem in which the errors have a Poisson-Normal distribution. Neither author, nor indeed Tibshirani, mentions robustness. However, Boente, Martínez, and Salibián-Barrera (2017) do provide a numerical

procedure for robust backfitting in the GAM. We believe our paper provides a robustification of AVAS in the spirit suggested by Buja and Kass (1985). The application of our methods to robustifying ACE would follow directly from the work presented here.

Alternatives to AVAS include Ramsay (1988) who uses a monotone spline to estimate the response transformation, with the regression parameters estimated by least squares. An advantage is that the Jacobian of the response transformation is found by straightforward differentiation of the spline. Ramsey also uses monotone splines for transformation of the explanatory variables. Breiman (1988) queries the restriction to monotonicity for the transformation of the explanatory variables, but commends the use of splines. Subroutine *transace* of the R-package `Hmisc` (Harrell Jr 2019) replaces the supersmoother in ACE and AVAS with restricted cubic smoothing splines, with a controllable number of knots.

The desirable property of AVAS, and of our robust version, is that they provide the flexibility of nonparametric models for both the response and the explanatory variables. Most of the large literature on generalized linear models (GLMs) and GAMs uses smoothing or spline methods on either the response or the explanatory variables, but not both. An exception is Spiegel, Kneib1, and Otto-Sobotka (2019) who use spline smoothing both for the link function in a GLM and for the transformation of the explanatory variables in the GAM. Robustness is introduced to fitting the GAM by Alimadad and Salibian-Barrera (2011) who use a soft-trimming method, combined with spline smoothing. Their results extend those of Cantoni and Ronchetti (2001) for the robust fitting of GLMs. Boente, Martínez, and Salibián-Barrera (2017) formulate the numerical procedure for robust backfitting in the GAM using polynomial smoothers.

In this article we present a robust and improved version of Tibshirani's algorithm. Our starting point was the original version of his algorithm, written in "classical" Fortran, that is with few comments and uninformative variable names. We first converted this nonrobust algorithm into Matlab. Our extension of AVAS is fully documented and freely available on GitHub. Details are given in the supplementary material. It is programmed to allow the use of a variety of methods of robust regression. These include S-estimation (Rousseeuw and Yohai 1984) and MM-estimation (Yohai and Zamar 1988), in which outliers are downweighted, and the hard trimming methods least Trimmed Squares (LTS) and the Forward Search (see Section 4.1.3) in which outliers have weight zero. In our numerical examples we use the Forward Search (Atkinson, Riani, and Cerioli 2010) to provide a subset of the observations believed to be outlier free. Because of the potential for an appreciable number of outliers which also need to be transformed for further calculations, we develop a new procedure for the necessary interpolation and extrapolation for these observations. Since the whole procedure involves iteration between response and explanatory variable transformation, the subset of observations treated as outlying may change from iteration to iteration. The backfitting and smoothing procedures for transforming the explanatory variables are estimated using this subset. Likewise, the numerical procedure for the variance stabilizing transformation of the response is only estimated from this subset, although it is applied to all observations.

In summary, the main purposes of our article, in order of increasing generality, are:

1. To clarify the details of Tibshirani's algorithm;

2. To provide a version with improved numerical estimation of variance used in constructing the transformation. Other improvements, introduced in Section 4, are also available as options;

3. To present a flexible robust version of response and explanatory variable transformations, with graphical interpretation which can be interactive;

4. Finally, to make these procedures freely available on GitHub.

In addition to robustness we have made four improvements to the performance of AVAS. These, like robustness, are available as options and so can be employed whether or not a robust option is chosen. The program has been written to be highly flexible and can, for example, be run solely for nonparametric response transformation in regression using least squares.

In the next section we provide some background to the AVAS algorithm described in Section 3. Our robust algorithm, RAVAS, is presented in Section 4. Robust regression is introduced in Section 4.1.3, followed by details of the variance stabilizing transformation.

The remainder of the article illustrates the use of RAVAS. In Section 5 we demonstrate the program's ability to detect outliers and illustrate the effect of our improvements in the algorithm on a nonrobust analysis. Section 6 introduces a novel graphical display (the augmented star plot) to illustrate the effectiveness of our five options on data analyses. It can also be brushed to link to the plot for particular analyses. These properties are illustrated on simulated data. In Section 7 we analyze data on the effectiveness of advertisements using social media. We conclude in Section 8 with brief details of a simulation study comparing RAVAS with traditional nonrobust AVAS, for samples of up to 10,000 observations. Unlike AVAS, RAVAS provided parameter estimates with mean squared error that was virtually constant and tests for outliers, the power of which tended to one as the outliers became more remote. Except for small sample sizes and moderate outliers, RAVAS converged in three iterations, whereas convergence of AVAS required increasingly many iterations as the outliers became more remote.

2. Background

2.1. Introduction

The generalized additive model (GAM) has the form

$$g(Y_i) = \beta_0 + \sum_{j=1}^{p} f_j(X_{ij}) + \epsilon. \tag{1}$$

The functions f_j are unknown and are, in general, found by the use of smoothing techniques. A monotonicity constraint can be applied. If the response transformation or link function

g is unknown, it is restricted to be monotonic, but scaled to satisfy the technically necessary constraint that $\text{var}\{g(Y)\} = 1$ (otherwise, the zero transformation would be trivially perfect). In the fitting algorithm the transformed responses are scaled to have mean zero; the constant β_0 can therefore be ignored. The observational errors ϵ are assumed to be independent and additive with constant variance. The performance of fitted models is compared by use of the coefficient of determination R^2. Since the f_j are estimated from the data, the traditional assumption of linearity in the explanatory variables is avoided. However, the GAM retains the assumption that explanatory variable effects are additive. Buja, Hastie, and Tibshirani (1989) describe the background and early development of this model.

In the next section we assume that the response transformation $g(y)$ is known and describe the backfitting algorithm for estimating the functions f_j.

2.2. Backfitting

2.2.1. Definition

The backfitting algorithm, described in Hastie and Tibshirani (1990, p. 91), is used to fit a GAM. The algorithm proceeds iteratively using residuals when one explanatory variable in turn is dropped from the model. A discussion of convergence of iterative versions is in Schimek and Turlach (2006).

With $g(y)$ the $n \times 1$ vector of transformed responses, let $e_{(j)}$ be the vector of residuals when $f_j(x_j)$ is removed from the model without any refitting. Then

$$e_{(j)} = g(y) - \sum_{k \neq j=1}^{p} f_k(x_k). \tag{2}$$

The new value of $f_j(.)$ depends on ordered values of $e_{(j)}$ and x_j. Let the ordered values of x_j be $x_{s,j}$. The residuals $e_{(j)}$ are sorted in the same way to give the new order $e_{s,(j)}$. Within each iteration each explanatory variable is dropped in turn; $j = 1, \ldots, p$. The iterations continue until the change in the value of R^2 is less than a specified tolerance.

For iteration l the vector of sorted residuals for x_j is $e_{(j)}^l$. The new estimate of $f_j^{(l+1)}$ is

$$f_{s,j}^{(l+1)} = S\left\{e_{s,(j)}^l, x_{s,j}\right\}. \tag{3}$$

The required estimates of $f_j^{(l+1)}$ follow by restoring the elements of $f_{s,j}^{(l+1)}$ to the appropriate values of unordered X_j.

The function S depends on the constraint imposed on the transformation of variable j.

- If the transformation can be non-monotonic, S denotes a smoothing procedure. In order to compare our procedure with that of Tibshirani (1988) we follow him in using the supersmoother (Friedman and Stuetzle 1982), a nonparametric estimator based on local linear regression with adaptive bandwidths.
- If the transformation is monotonic, the $f_j^{(l+1)}$ come from isotonic regression (Barlow et al. 1972).
- If the transformation is linear $f_j^{(l+1)} = a + bX_j$, where a and b are the least squares coefficients.

2.2.2. Properties

For a linear regression model with errors that obey second-order assumptions, the least squares estimates of the parameters satisfy the normal equations. The backfitting algorithm is the Gauss-Seidel algorithm for solving this set of equations. Buja, Hastie, and Tibshirani (1989) extend this result for least squares to the backfitting algorithm for linear smoothers; a smoother is linear if $\hat{y} = Sy$ and S does not depend on y. This result assumes that, if necessary, the response has been transformed so that the second-order conditions are satisfied. Unfortunately, the supersmoother that is used for transformation of explanatory variables is not linear, so that convergence of the procedure is not automatically guaranteed.

The backfitting algorithm is not invariant to the permutation of order of the variables inside matrix X, with high collinearity between the explanatory variables causing slow convergence of the algorithm: the residual sum of squares can change very little between iterations (Breiman and Friedman 1985; Hastie and Tibshirani 1988). Our option orderR2, Section 4.1.4, attempts a solution to this problem by reordering the variables in order of importance.

3. Introduction to AVAS

In this section we first present the structure of the AVAS algorithm of Tibshirani (1988) and then outline the variance stabilizing transformation used to estimate the response transformation. Our RAVAS algorithm has a similar structure, made more elaborate by the requirements of robustness and the presence of options.

3.1. The AVAS Algorithm

In this description of the algorithm ty and tX are transformed values of y and X. In both algorithms the inner loop fits the GAM for a given $g(y)$ leading to new values of tX and residuals e. The main (outer) loop uses these to calculate a new transformation $g(y)$.

1. *Initialize Data.* Standardize response y so that $\overline{ty} = 0$ and $\text{var}(ty) = 1$, where var is maximum likelihood biased estimator of variance. Center each column of the X matrix so that $\overline{tX}_j = 0, j = 1, \ldots, p$).

2. *Initial call to "Inner Loop"* to find initial GAM using ty and tX; calculates initial value of R^2. Set convergence conditions on number of iterations and value of R^2.

3. *Main (Outer) Loop.* Given values of ty and tX at each iteration, the outer loop provides updated values of the transformed response. Given the newly transformed response, updated transformed explanatory variables are found through the call to the backfitting algorithm (*inner loop*). In our version iterations continue until a stopping condition on R^2 is verified or until a maximum number of iterations has been reached.

3.2. The Numerical Variance Stabilizing Transformation

We first consider the case of a random variable Y with known distribution for which $\text{E}(Y) = \mu$ and $\text{var}(Y) = V(\mu)$. We seek

a transformation $ty = h(y)$ for which the variance is, at least approximately, independent of the mean. Then Taylor series expansion of $h(y)$ leads to $\text{var}(Y) \approx V(\mu)\{h'(\mu)\}^2$. For a general distribution $h(y)$ is then a solution of the differential equation $dg/d\mu = C/\sqrt{V(\mu)}$. For random variables standardized, as are the values ty, to have unit variance, $C = 1$ and the variance stabilizing transformation is

$$h(t) = \int^t 1/\sqrt{V(u)}du. \qquad (4)$$

In the AVAS algorithm for data, $1/\sqrt{V(u)}$ is estimated by the vector of the reciprocals of the absolute values of the smoothed residuals sorted using the ordering based on fitted values of the model. There are n integrals, one for each observation. The range of integration goes from the smallest fitted value, to $\widehat{ty}_i, i = , \dots, n$. The computation of the n integrals uses the trapezoidal rule. The logged residuals in the estimation of the variance function are smoothed using the running line smoother of Hastie and Tibshirani (1986). Full details of our implementation of the algorithm are in Atkinson et al. (2023, chap. 7).

4. RAVAS: Five Extensions to AVAS

Our RAVAS procedure introduces five improvements to AVAS, programmed as options. These do not have a hierarchical structure, so that there are 2^5 possible choices of the options. The augmented star plot of Section 6 provides a method for assessing these choices. We discuss the motivation and implementation for each in the order in which they are applied to the data when all are employed.

4.1. Initial Calculations

The structure of our algorithm is an elaboration of that of AVAS outlined in Section 3.1. Four of the five options can be invoked before the start of the outer loop. They are described here. A detailed flow scheme is in Section 2 of the supplementary material.

4.1.1. Initialization of Data: Option tyinitial
Our numerical experience is that it is often beneficial to start from a parametric transformation of the response. This is optionally found using the automatic robust procedure for power transformations described by Riani, Atkinson, and Corbellini (2023). For $\min(y) > 0$ we use the Box-Cox transformation. For $\min(y) \leq 0$ the extended Yeo-Johnson transformation is used (Atkinson, Riani, and Corbellini 2020). This family of transformations has separate Box and Cox transformations for positive and negative observations. In both cases the initial parametric transformations are only useful approximations, found by searching over a coarse grid of parameter values. The final nonparametric transformations sometimes suggest a generalization of the parametric ones.

4.1.2. Ordering Explanatory Variables in Backfitting: Option scail
In a comparison of monotone regression spline estimation with ACE, Ramsay (1988) observed that the fitted model obtained with ACE depends on the order of the explanatory variables (see Section 2.2). One approach is to use an initial regression to remove the effect of the order of the variables through scaling (Breiman 1988). With b_j the coefficient of $f_j(x)$ in the multiple regression of $g(y)$ on all $f_j(x)$ the option scail provides new transformed values for the explanatory variables: $\widehat{tX}_j = b_j f_j(x)$, $j = 1, \dots, p$. Option scail is used only in the initialization of the data.

4.1.3. Robust Regression and Robust Outlier Detection: Option rob
1. Robust Regression.
We robustify our transformation method through the use of robust regression to replace least squares. There are three main approaches to robust parametric regression.

1. Hard (0,1) Trimming. In Least Trimmed Squares, Hampel (1975), Rousseeuw (1984) the amount of trimming is determined by the choice of the trimming parameter h, $[n/2] + [(p + 1)/2] \leq h \leq n$, which is specified in advance. The LTS estimate is intended to minimize the sum of squares of the residuals of h observations. For least squares (LS), $h = n$.

2. Adaptive Hard Trimming. In the Forward Search (FS), the observations are again hard trimmed, but the value of h is determined by the data, being found adaptively by the search. Atkinson, Riani, and Cerioli (2010) provide a general survey of the FS, with discussion.

3. Soft trimming (downweighting). M estimation and derived methods. The intention is that observations near the center of the distribution retain their value. Increasingly remote observations are downweighted by a ρ function to give a weight that decreases with distance from the fitted model. We include S and MM estimation. A popular ρ function is Tukey's biweight (Beaton and Tukey 1974).

All three methods provide weights for the observations. Our RAVAS algorithm has been programmed to use any of the three robust methods. However, we prefer the Forward Search for outlier detection, since it depends upon the data to determine the number and identity of the outliers, rather than analyzing the data with the pre-determined value for h of LTS. All examples in this article have been computed using the FS. Comparisons of outlier detection using hard and soft-trimming in regression are in Riani, Atkinson, and Perrotta (2014).

2. Robust Outlier Detection
Our algorithm works with k observations treated as outliers, providing the subset of $m = n - k$ observations used in model fitting and parameter estimation. This section describes our outlier detection methods.

The default setting of the forward search uses the multivariate procedure of Riani, Atkinson, and Cerioli (2009) adapted for regression (Torti, Corbellini, and Atkinson 2021) to detect outliers at a simultaneous level of approximately 1% for samples of size up to around 1000. Optionally, a different level can be selected.

In the other two methods of robust regression we again calculate scaled residuals for all n observations and use a Bonferroni inequality to give a simultaneous test for outliers with significance level for detection of individual outliers of α/n. The m observations for which the test is not significant then form the subset S_m.

We use the subset S_m in the backfitting algorithm to calculate the transformations $f(X)$ at each iteration, the k outliers being ignored as they are in the calculation of the numerical variance stabilizing transformation in Section 4.2.1. Since different response transformations can indicate different observations as outliers, the identification of outliers occurs repeatedly during our robust algorithm, once per iteration of the outer loop. Examples of the dependence of outliers on the parameter of the Box-Cox transformation are in Atkinson and Riani (2000, chap. 4). Here there is also a relationship between the transformations $f(x)$ and the declaration of outliers.

4.1.4. Ordering Predictor Variables: Option orderR2

In order to completely eliminate dependence on the order of the variables, we include an option that, at each iteration, provides an ordering which is based on the variable which produces the highest increment of R^2. Let $A_{j-1} = \{i_1, \ldots, i_{j-1}\}$ be the set of the $j-1$ indexes for the columns of matrix X which have already been updated, and let i_r be the index of a column of X that has not yet been updated; then $i_r \notin A_{j-1}$. Further, let $R^2(i_r)$ be the coefficient of determination in a multiple linear regression which has g as response and as regressors $f_{i_1}, \ldots, f_{i_{j-1}}, f_{i_r}$. We select as the next variable i_s to be estimated in the backfitting algorithm, that for which

$$i_s = \arg \max_{i_r \notin A_{j-1}} R^2(i_r).$$

With this criterion the most relevant features are immediately transformed and those that are perhaps irrelevant will be transformed in the final positions. For robust estimation, this procedure is applied solely to the observations in the subset S_m. Option orderR2 is available at each call to the backfitting function. Flowcharts for initialization and iterative use of the outer loop of RAVAS are in the supplementary material.

This robust procedure, with ordering, offers interesting suggestions for robust model selection for GAMs. We do not here develop this idea further.

4.2. Main or Outer Loop

The only option remaining to be described is the robust version of the numerical variance stabilizing transformation of Section 4.2.1.

4.2.1. Numerical Variance stabilizing Transformation: Option Trapezoid

Plots of residuals against fitted values are widely used in regression analysis to check the assumption of constant variance (Anscombe and Tukey 1963; Cox and Snell 1981). Here the observations have been transformed, so the fitted values are $\widehat{ty}_i$. To estimate the variance stabilizing transformation, the fitted values have to be sorted, giving a vector of ordered values $\widehat{ty}^s$. The residuals are ordered in the same way and, following the

procedure of Section 3.2, provide estimates v_i of the integrand $V^{-0.5}(y)$ in (4). The v_i provide estimates at the ordered points $\widehat{ty}^s$. Calculation of the variance transformation (4) is however for sorted observed responses ty_i^s, rather than fitted, transformed responses $\widehat{ty}^s$. As did Tibshirani, we use the trapezoidal rule to approximate the integral. Linear interpolation and extrapolation are used in calculation of the v_i at the ty_i^s. We provide an option "trapezoid" for the choice between two methods for the extrapolation of the variance function estimate, the interpolation method remaining unchanged. Our approach leads to trapezoidal summands in the approximation to the integral for the extrapolated elements, whereas Tibshirani's leads to rectangular elements. When we are concerned with robust inference, there are only $m = n - k$ members of $\widehat{ty}^s$ whereas there are n values of ty_i^s, so that robustness increases the effect of the difference between the two rules. The option trapezoid = false uses rectangular elements in extrapolation.

For a transformation that stabilizes variance, there is no relationship between var(ty) and ty. Thus, the smoothed values of the residuals will be constant, apart from random fluctuations.

A more detailed discussion of the calculation of the variance stabilizing transformation, together with the flowchart, are in the supplementary materials. A figure and clarificatory discussion are in Atkinson et al. (2023, chap. 7).

5. Simulated Examples

In this section we start with a simple example to compare the data analyses from RAVAS with the nonrobust AVAS of Tibshirani when there are a few outliers. In Section 5.2 our analysis is of data without introduced outliers to demonstrate the effect of our options on the performance of the nonrobust algorithm.
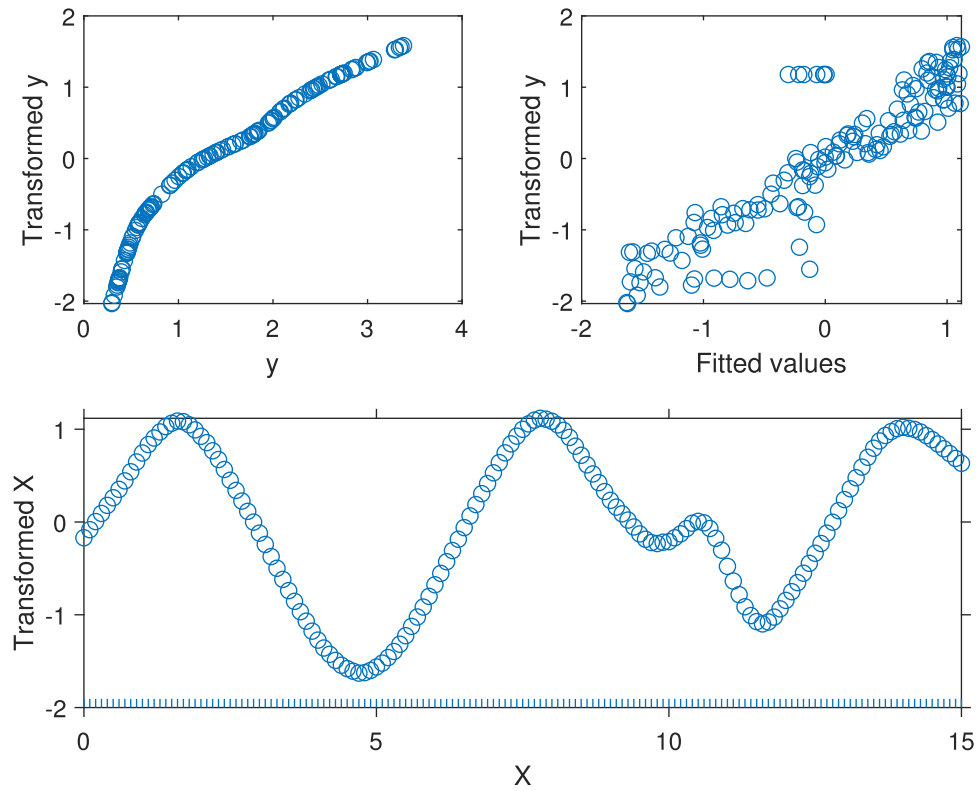
5.1. Example 1—Robustness

There are 151 observations with x equally spaced over 0,(0.1),15. The linear model is

$$z = \sin(x) + 0.5(\mathcal{U}(0, 1) - 0.5).$$

Outliers of value one replace the values of z in positions 100, 105, 106,…110. That is there are seven outliers at $x = 9.9, 10.4, 10.5, \ldots, 10.9$. The response $y = \exp z$. Thus, a logarithmic transformation of the observed responses is appropriate. We start with a nonrobust analysis excluding our improvements to the algorithm. That is, there is no initial transformation of the response (tyinitial = false) and we use the rectangular extrapolation when calculating the variance stabilizing transformation (trapezoid = false).

The top left-hand panel of Figure 1 shows the transformation of y, which is roughly logarithmic, but with a slight decrease in gradient near the centre of the plot. The top right-hand panel shows the transformed values of y against fitted values. Two horizontal linear structures are apparent; a set of six values of transformed y close to one and several values close to -1.75. The bottom panel of Figure 1 shows a very clear distortion of the sinusoid of $f(x)$ in the range of the outliers, that is 9.9–10.9. This "rug" plot shows the uniform distribution of the values of x.

In our robust analysis we used the initial transformation of y and set trapezoid = true. The top panels of Figure 2 show plots of
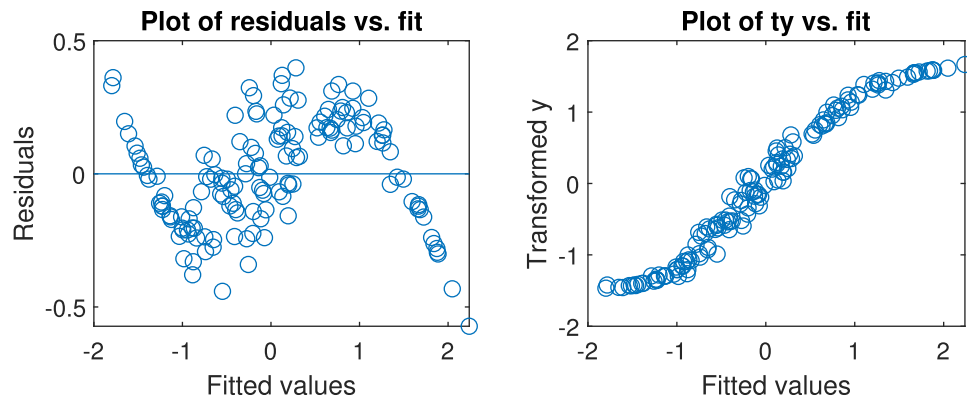
Figure 1. Example 1 (seven outliers). Standard nonrobust analysis without options. Top left-hand panel, transformed *y* against *y*; top right-hand panel, transformed *y* against fitted values; lower panel, transformed *x* against *x*.
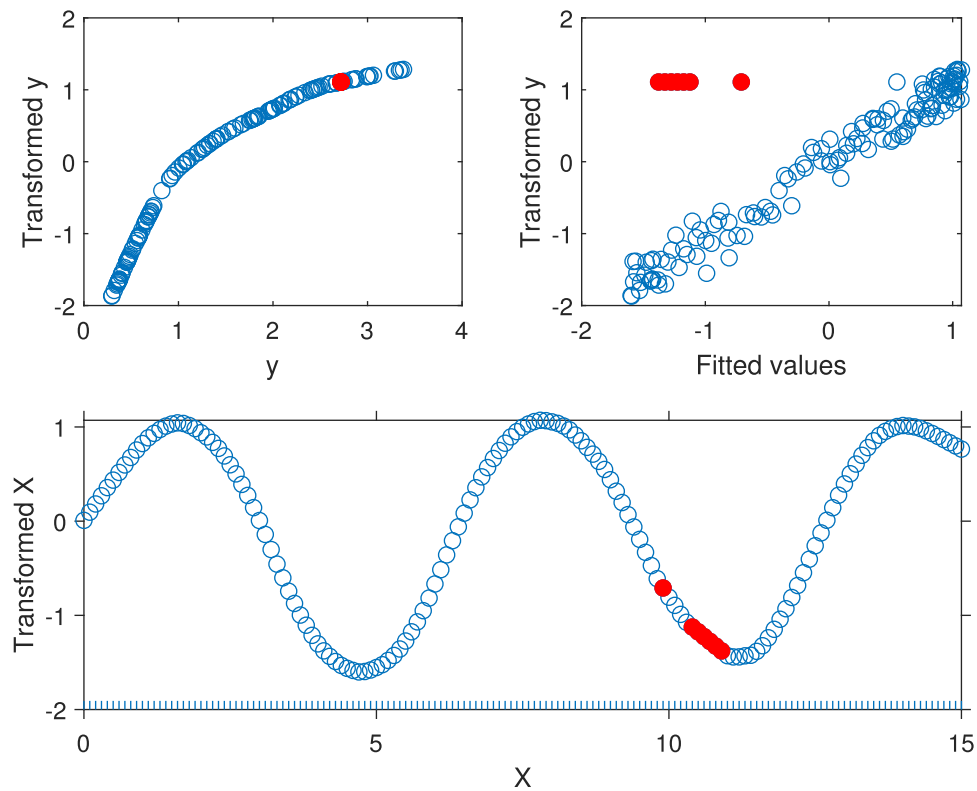
Figure 2. Example 1. Robust analysis. Top left-hand panel, transformed *y* against *y*; top right-hand panel, transformed *y* against fitted values; lower panel, transformed *x* against *x*. The seven outliers are indicated by filed symbols.

the transformed values of *y* with the seven outliers automatically detected by the robust procedure. The plot against *y* shows a smoother curve than that for the nonrobust analysis. However, the greater difference is in the plot against fitted values. There is now a regression line with less scatter than in Figure 1; the linear structure at the bottom of the plot is now absent and the

Figure 3. Two-variable model. Nonrobust analysis without options. Left-hand panel, residuals against fitted values; right-hand panel, transformed *y* against fitted values.

horizontal line formed by the group of six outliers is more clearly distanced from the fitted regression. The plot also shows the isolated outlier (corresponding to $x = 9.9$). The bottom panel of Figure 2 shows that the robust analysis has removed the distortion in the estimation of the sinusoidal form of $f(x)$, except where the outliers have been deleted.

In this and other figures the outliers are indicated by filled symbols. In the online .pdf version they are colored red.

5.2. Example of Effects from Options in RAVAS

The preceding example shows that the robustness of RAVAS leads to the identification of the seven outliers and to the virtual recovery of the model from which the data were simulated. In this section we use further simulations to explore how our modifications, not necessarily the robustness itself, lead to improved analyses of data.

Two-variable model. We use an example without outliers to illustrate the importance of some options including the initial transformation of the response. There are 151 observations with $x_1 \sim N(0, 1)$. The linear model is

$$z = 10\{1 + \sin(x_1) + \exp(x_2)\} + \mathcal{N}(-0.5, 0.5),$$

where $x_2 = 0, (0.01), 1.5$. The observed responses are $y = z^3$, so that a 1/3rd transformation is expected. As well as plotting residuals and fitted values from some of the fitted models, we also calculate the value of R^2 and the significance level of the Durbin-Watson statistic (Durbin and Watson 1950) for the independence of the residuals ordered by the fitted values from the model.

Figure 3 shows results from using RAVAS on these data without any options. The left-hand panel shows the plot of residuals against fitted values which has a sinusoidal shape. The plot of transformed y against fitted values in the right-hand panel is logistic in shape, rather than linear. These two plots indicate that the fitted model is not satisfactory, an impression confirmed by the value of 2.3913×10^{-21} for the significance of the Durbin-Watson statistic. We do not show the plot of transformed y against y which is roughly appropriate for a transformation of 1/3.

We now proceed adding options in an ad hoc manner, starting with scail, that is the rescaling of the explanatory variables, Section 4.1.2. Figure 4 shows plots of the same properties as

Figure 3. The plot of the residuals now shows a series of rough diagonal bands with an outline for larger fitted values similar to that in Figure 3. Although straighter, the plot of transformed against fitted y is again curved for higher values of y. These improvements lead to a less significant value of 2.0349×10^{-14} for the significance of the Durbin-Watson statistic. The fit is still far from satisfactory.

In the modeling producing the third pair of figures (Figure 5) we have also included the initial transformation of the responses (tyinitial, Section 4.1.1). It is clear from scatterplots of the data that the transformation parameter λ should be positive but small. We used a grid of $\lambda = 0, 0.1, 0.2, 0.3, 0.4,$ and 0.5.
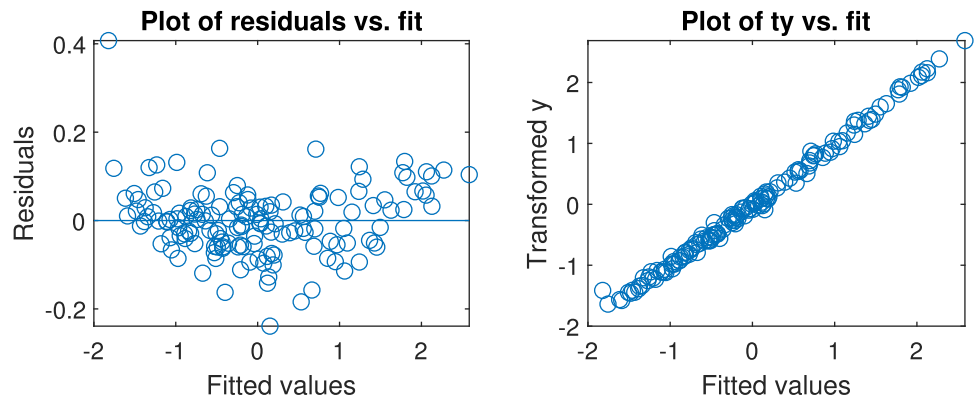
The plot of residuals against fitted values in the left-hand panel of the figure is greatly improved from that of Figure 4 but is still not random, showing some curvature. This is reflected in the improved, but still significantly small, value of 3.225×10^{-7} for the significance of the Durbin-Watson test. The plot of transformed y against fitted values is closer to a straight line than before.

These three different sets of options show steady improvement in the normality of the residuals. We have examined them in some detail because we observed plots with structures similar to those in the left-hand panels of Figures 3 and 4 when we used AVAS to analyze data for response transformations in Atkinson, Riani, and Corbellini (2020), although we did not publish such plots. We have also failed to find any references to such phenomena. We conclude this progression with the analysis using all five options.

The upper left-hand panel of Figure 6 shows the smooth curve of the transformed responses and the upper right-hand panel shows the plot of residuals against fitted values. This appears to be a random scatter with, surprisingly, two outliers, an unexpected artifact of the simulation. The significance value for the Durbin-Watson statistic is 0.77, so that independence of the residuals has been achieved, once the outliers are deleted. The bottom left-hand panel shows the linear relationship between transformed response and fitted values, as is expected for a GAM, together with the two outliers. The estimated transformations of the two explanatory variables are in the bottom right-hand panel of the figure. The first explanatory variable shows the shape of a scaled and translated sine function, with the two observations for the largest values of x_1 being marked as outlying. The second explanatory variable has the desired exponential form. Plots of these nonparametric functions with
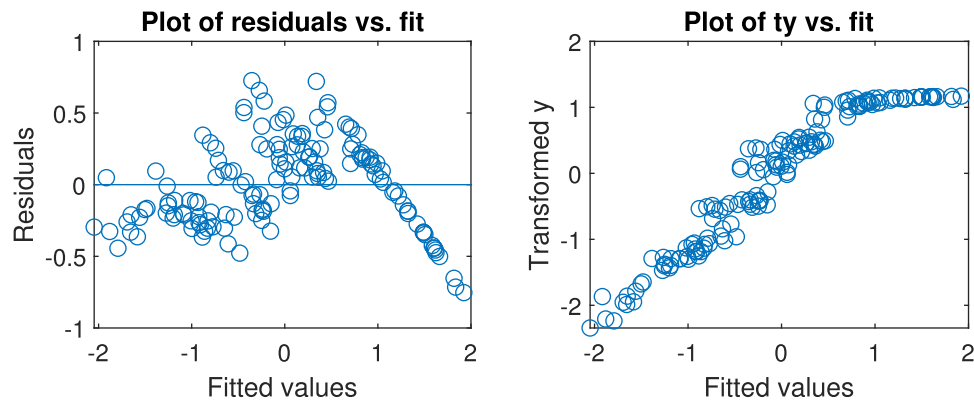
Figure 4. Two-variable model. Nonrobust analysis with option scail. Left-hand panel, residuals against fitted values; right-hand panel, transformed *y* against fitted values.

Figure 5. Two-variable model. Nonrobust analysis with options scail and tyinitial. Left-hand panel, residuals against fitted values; right-hand panel, transformed *y* against fitted values.
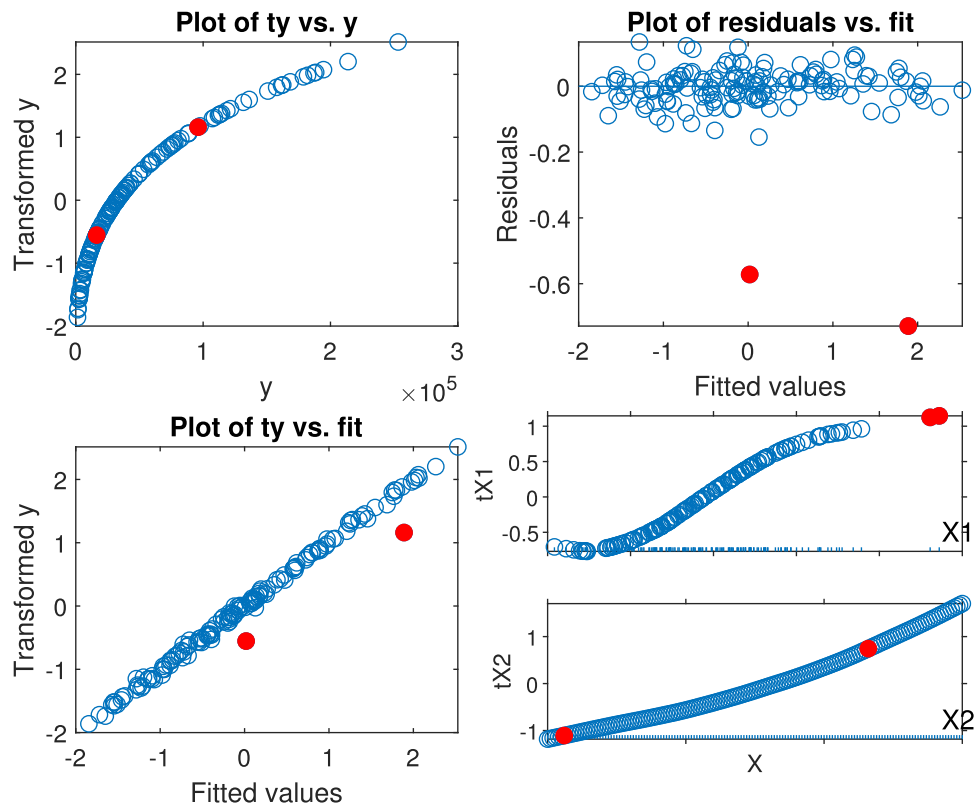
Figure 6. Two-variable model. Robust analysis with all options. Upper left-hand panel, transformed *y* against *y*; upper right-hand panel, residuals against fitted values; lower left-hand panel, transformed *y* against fitted values; lower right-hand panel, transformed explanatory variables. Two outliers are shown by filled symbols
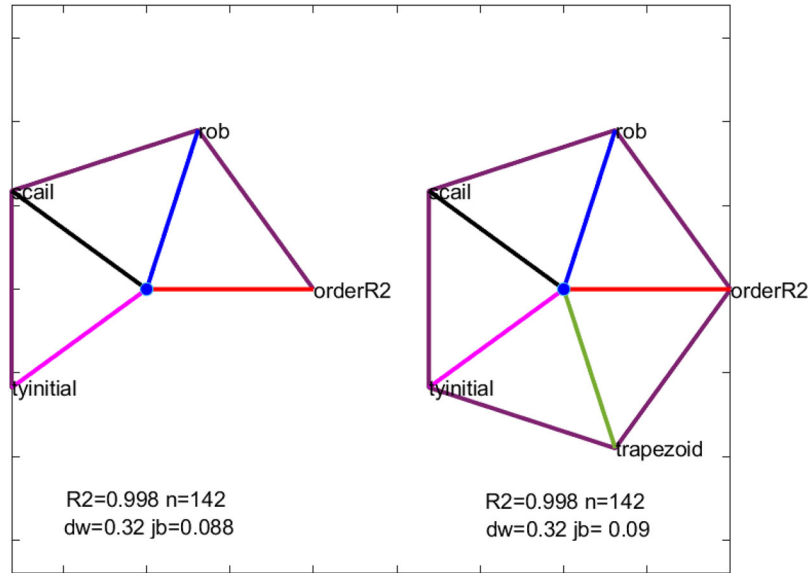
Figure 7. Example 2. Augmented star plot of options. The best solution has four options and the statistically indistinguishable second-best solution includes the trapezoid option.

superimpositions of the parametric forms can be used to check closeness to the parametric models.

6. A Graphical Representation of the Systematic Evaluation of Options

We have added five options to the original AVAS. There are therefore 32 combinations of options that could be chosen. It is not obvious that all will be necessary when analyzing all sets of data. Our program provides flexibility in the assessment of these options. One possibility is a list of options ordered by, for example, the value of R^2 or of the significance of the Durbin-Watson or the normality test. In this section we describe the augmented star plot, one graphical method for visualizing interesting combinations of options in a particular data analysis. An example is Figure 7. Figure 10 introduces a brushable version of the augmented star plot which links to residual and other plots from the fitted model.

We remove all analyses for which the residuals fail the Durbin-Watson and Jarque-Bera normality tests, at the 10% level (two-sided for Durbin-Watson). The Jarque-Bera (Jarque and Bera 1987) test uses a combination of estimated skewness and kurtosis to test the distributional shape of the residuals. There is an option for independent choice of the levels of these two tests. We used a threshold of 5% for both. We optionally enclose them in a polygon.

We order the remaining, admissible, solutions by the Durbin-Watson significance level multiplied by the value of R^2 and by the number of units not declared as outliers. Other options are available. The lengths of rays in individual panels of the plot are of equal length for those features used in an analysis; we enclose them in a polygon. All rays are in identical places in each panel of the plot; the lengths of the rays for each analysis are proportional

to p_{DW}, the significance level of the Durbin-Watson test. Each ray has a different color.

The ordering in which the five options are displayed in the plot depends on the frequency of their presence in the set of admissible solutions. For example, if robustness is the one which has the highest frequency, its ray is shown on the right. The remaining options are displayed counterclockwise, in order of frequency.

We list the five options giving an informative text description followed by the optional input arguments.

1. Initial robust transformation of the response; tyinitial—Section 4.1.1.

2. Remove effect of order of explanatory variables by regression; scail—Section 4.1.2.

3. Robustness; rob—Section 4.1.3.

4. Ordering variables in the backfitting algorithm using R^2 values; orderR2—Section 4.1.4.

5. Trapezoidal or rectangular rule for numerical integration; trapezoid—Section 4.2.1.

Example 2. As an initial example we consider an extension of Example 1 simulated in Section 5.1 in which there are now four explanatory variables. There are again 151 observations with x_1 equally spaced from 0,(0.1),15. The linear model is

$$z = \sin(x_1) + \sum_{j=2}^{4}(j-1)x_j + 0.5\mathcal{U}(-0.5, 0.5),$$

where $x_{ij}(j = 2, \ldots, 4)$, are independently $\mathcal{N}(0, 1)$. Eight outliers of value one replace the values of z at $x = 8.9, 9.9, 10.4,$
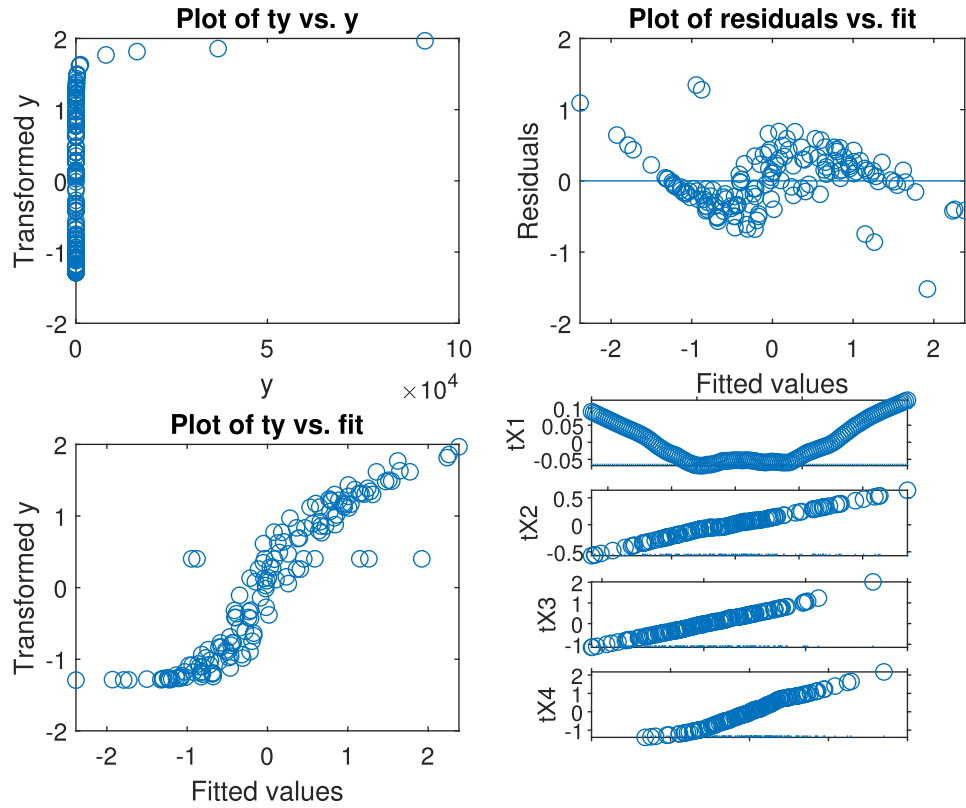
Figure 8. Example 2. Nonrobust analysis. Upper left-hand panel, transformed *y* against *y*; upper right-hand panel, residuals against fitted values; lower left-hand panel, transformed *y* against fitted values; lower right-hand panel, transformed explanatory variables.

10.5 ... 10.9. The response $y = \exp z$. There are thus four explanatory variables, only one of which requires transformation, as does the response to $\log y$.

Figure 7 shows the augmented star plots of the two combinations of options for which the residuals pass the tests of independence and normality. The first solution uses four options with the trapezoidal integration rule added for the second solution. The statistical properties of the two solutions are indistinguishable. For both, $R^2 = 0.998$, $p_{DW} = 0.32$ and p_{JB} is 0.088 and 0.090.

We now compare the analysis without options with that suggested in Figure 7. Figure 8 shows the nonrobust results from using AVAS. The value of $R^2 = 0.884$. Although the response has been adequately transformed the residuals are far from random (the significance level for the Durbin-Watson test is 1.64×10^{-8}, the sinusoid for x_1 is distorted and x_4 has been subjected to nonlinear transformation.

The fit from RAVAS using all options identifies nine outliers, the eight generated intentionally and one at unit 53. The upper-right and lower-left panels of Figure 9 show, after the outliers have been deleted, that the residuals are structure free when plotted against fitted values and that there is a linear relationship between the transformed values of y and the fitted values. The lower right-hand panel plots the transformed values of x_1. The sinusoid has been recovered. We do not show the plots of the transformations of the other explanatory variables, which are straight lines, unlike some of the plots in the lower right-hand panel of Figure 8.

Example from Wang and Murphy. Wang and Murphy (2005) used ACEPACK (Spector et al. 2016) to illustrate the use of ACE

and AVAS in the transformation of data. Their illustrations did not include any robust methods. We now illustrate the use of RAVAS to analyze one of their examples to which we have introduced some outliers. We also introduce an extended version of the augmented star plot which has advantages when, as here, many statistical options provide satisfactory fits to the data. The extension also allows brushing.

There are 200 observations and four explanatory variables all independently uniformly distributed on $[-1, 1]$. The linear model is

$$z = 4 + \sin(3x_1) + \text{abs}(x_2) + x_3^2 + x_4 + N(0, 0.1) \quad (5)$$

and $y = \log z$. There are nine outliers each randomly generated as $1.9 + 0.01\mathcal{N}(9, 1)$.

Figure 10 shows the six combinations of options which satisfy the constraints on probabilities. As expected, these are plotted in a different order from those in Figure 7. Solution 4 uses only robustness, which is included in all six combinations of options. The optional bars in the figure, representing the four properties of the fitted models, have been introduced as being easier to interpret than the list of numerical values, such as those in Figure 7, when the figure displays several solutions. Since the maximum number of outliers is $[n/2]$ we plot $(n - 2k)/n$. Brushing the figure reveals plots such as those in Figures 11 and 12, depending on which augmented star plot is brushed. Details are in section of the supplementary material.

The values of R^2 and of p_{JB} do not change as much as those for p_{DW}. The correlations between the pairs of fitted values for the various combinations have a minimum value of 0.996. There is virtually no difference between the fitted models, despite
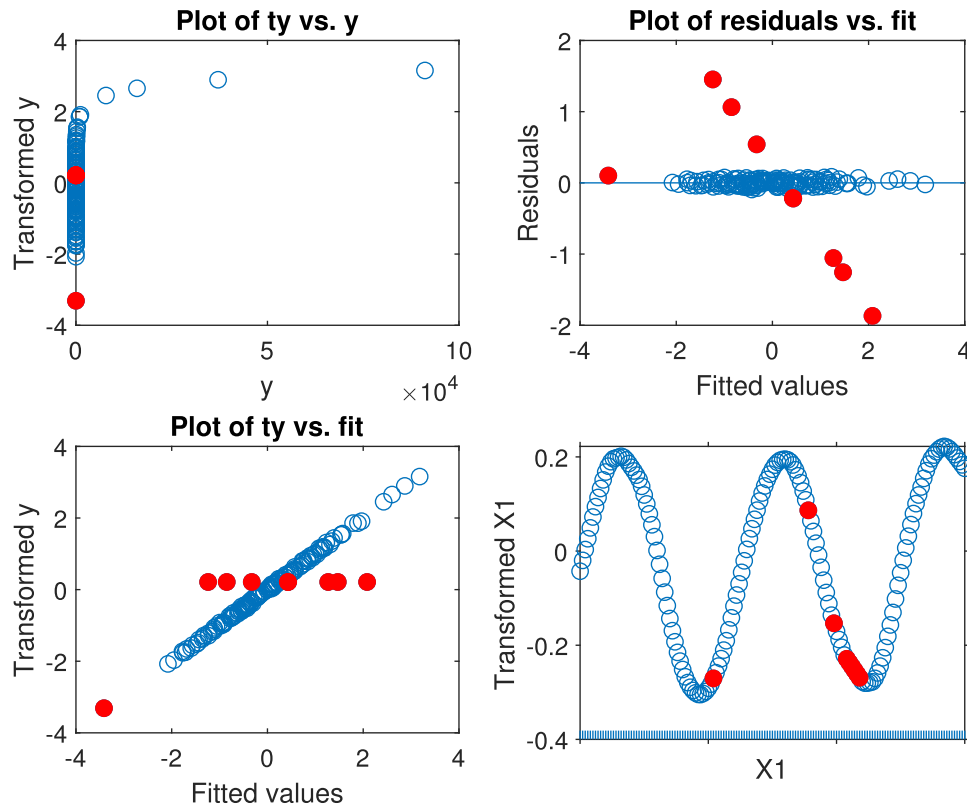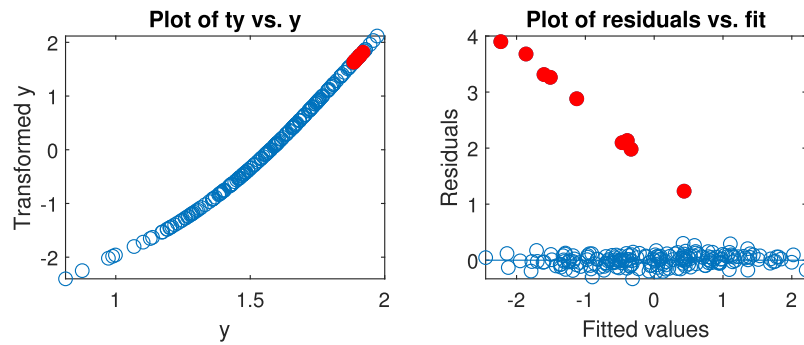
Figure 9. Example 2. Robust analysis with all options. Upper left-hand panel, transformed y against y; upper right-hand panel, residuals against fitted values; lower left-hand panel, transformed y against fitted values; lower right-hand panel, transformed explanatory variable x_1. Nine outliers shown by filled symbols.
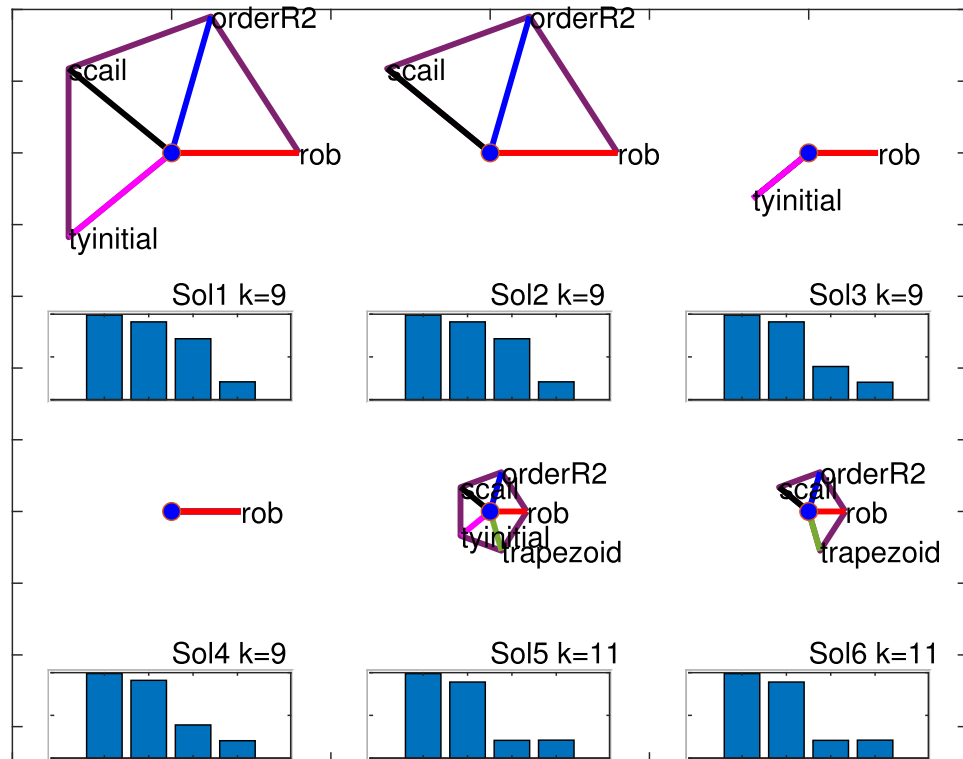
Figure 10. Example from Wang and Murphy. Extended augmented star plot of six selections of options. Solution 4 employs only robustness, which is included in all options. The solutions are statistically indistinguishable. The number of outliers detected is k. From left to right the bars give values of R^2, $(n - 2k)/n$, p_{DW} and p_{JB}.

the four selections of options detecting nine outliers and two detecting 11.

We look at some properties of the best solution from Figure 10, which uses all options apart from the trapezoidal rule.

Figure 11. Example from Wang and Murphy. Left-hand panel, transformed *y* against *y*; right-hand panel, residuals against fitted values. Nine outliers shown by filled symbols.

The plot of transformed *y* against *y* in the left-hand panel of Figure 11 shows the exponential relationship which is the inverse of the logarithmic transformation used to generate the data. The plot of residuals against fitted values in the right-hand panel shows the lack of relationship between the residuals and the fitted values. The nine outliers are clearly identified, although they are not evident in the scatterplots of the data. The transformations for the four explanatory variables (Figure 12) recover the model used in simulation (5). Particularly impressive is the plot for x_2 where the sharpness of the absolute value transformation is achieved by the smoothing algorithm. Similar results are obtained by Wang and Murphy (2005) who continue their analysis to the important step of testing the fit of parametric versions of these nonparametric forms for $f_j(x_j)$.

7. Internet Marketing Data

The R package datarium (Kassambara 2019) contains 200 results of an experiment on internet marketing. The explanatory variables $x_1 - x_3$ are the budgets for advertising on YouTube, Facebook, and newspapers and the response is sales. These data starkly show the gains in data analysis that come from using RAVAS with options rather than AVAS.

We start with linear regression on the three explanatory variables, which gives an *F* statistic for regression of 504 and a *t* statistic for x_3 of 0.658. That newspapers advertising has no effect and that there is a strong effect of advertising in the two online media is a constant pattern in all our analyses. We gain more insight into the structure of the data by using AVAS. It makes sense to assume that increasing advertising expenditures increase sales, so we include a monotonicity constraint on the transformations of the explanatory variables.

The resulting regression gives an *F*-value of 472 (even smaller than that for linear regression). Figure 13 shows that there is a strongly nonrandom curved pattern in the plot of residuals against fitted values, despite the smooth transformation of the response. Use of RAVAS with all options leads to the deletion of 17 observations giving an *F* value of 936. Despite the number of observations deleted, the residuals fail both tests of normality—with significance levels dw = 2×10^{-7} and jb = 0.0095.

Figure 14 shows the plot of residuals against fitted values in the RAVAS analysis with all options. This figure shows that the residuals are not yet free of structure. The deletion of the large number of outliers has reduced the curved pattern evident in

Figure 13. The deletion of 17 outliers has also produced relatively smooth curves in the other panels, except for the transformation of x_3 which has a bizarre stepped shape. Since this variable is not significant, the pattern can be ignored.

The curved pattern that remains in the plot of fitted values against residuals, ignoring outliers, suggests that a second-order model might be appropriate. We drop x_3 from the model and fit a full second-order model in x_1 and x_2, including quadratic and interaction terms. The augmented star plot, Figure 15, shows that the preferred model includes the options tyinitial, rob, and scail. The significance levels of the two tests are dw =0.98 and jb = 0.033. This fit detects one outlier and produces an *F*-statistic for regression of 5579, a more than 10-fold increase from our first fit using AVAS. The results in Figure 16 show a smooth transformation of the response and a linear relationship between the transformed response and fitted values, once the outlier is ignored. The transformed explanatory variables all have smoothly changing shapes and all are significant; the least significant *t*-statistic is −3.01.

To conclude our analysis we look at the results of fitting (nonrobust) AVAS with a quadratic model in two variables, that is, excluding x_3. The results, compared with RAVAS, which led to the deletion of one observation, are amazing. There are, of course, no deleted outliers. The *F*-statistic now has a value of 556, hardly better than regression on the first-order model without transformation of the response or of the explanatory variables. The plot (not given here) of *tXj* against *Xj* shows that there are smooth transformations of all five explanatory variables. However, the plot of the transformed response against *y* (left-hand panel of Figure 17) includes an unexpected kink. The plot of transformed *y* against fitted values (again not shown) is not a straight line, but is lumpy for low values. Most surprisingly, the plot of residuals against fitted values (right-hand panel of Figure 17), shows the inexplicable diagonal structures that we noticed in Figures 4 and 8. Such structures were the original stimulus for our attempts to provide a reliable version of AVAS that produces interpretable data analyses.

8. Discussion

Our examples in this article illustrate the excellent properties of our robust procedure. Even if robustness is not of interest, the results of Section 5.2 show how we improved on the performance of Tibshirani's algorithm. Harrell Jr (2019) claims that
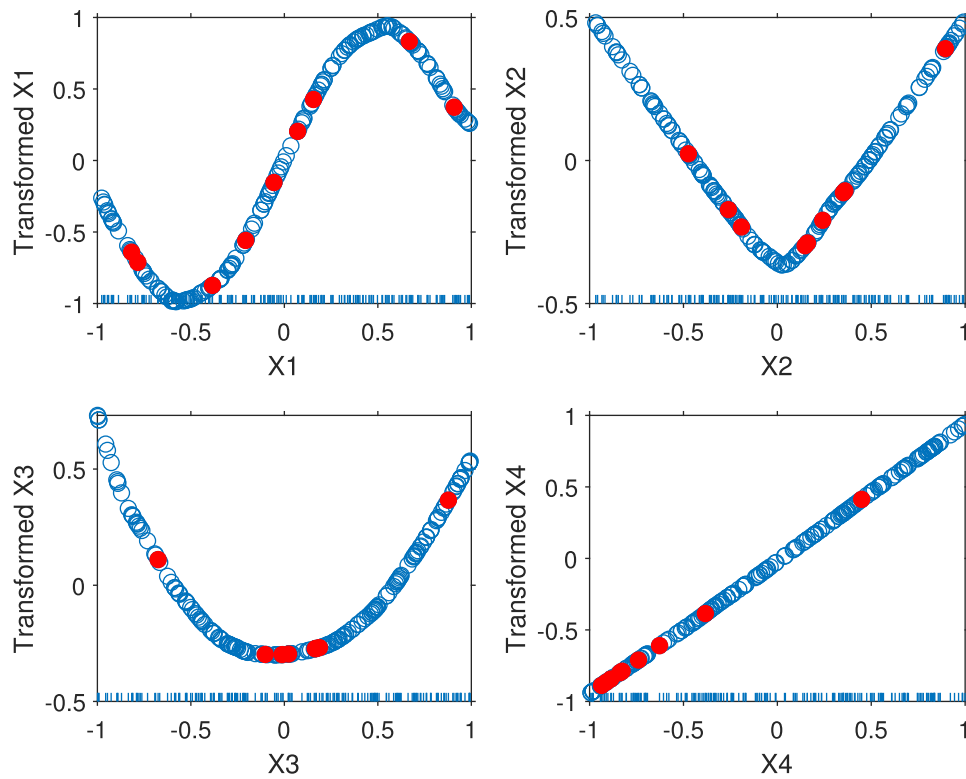
Figure 12. Example from Wang and Murphy. Transformations of the four explanatory variables. Nine outliers plotted as filled symbols.
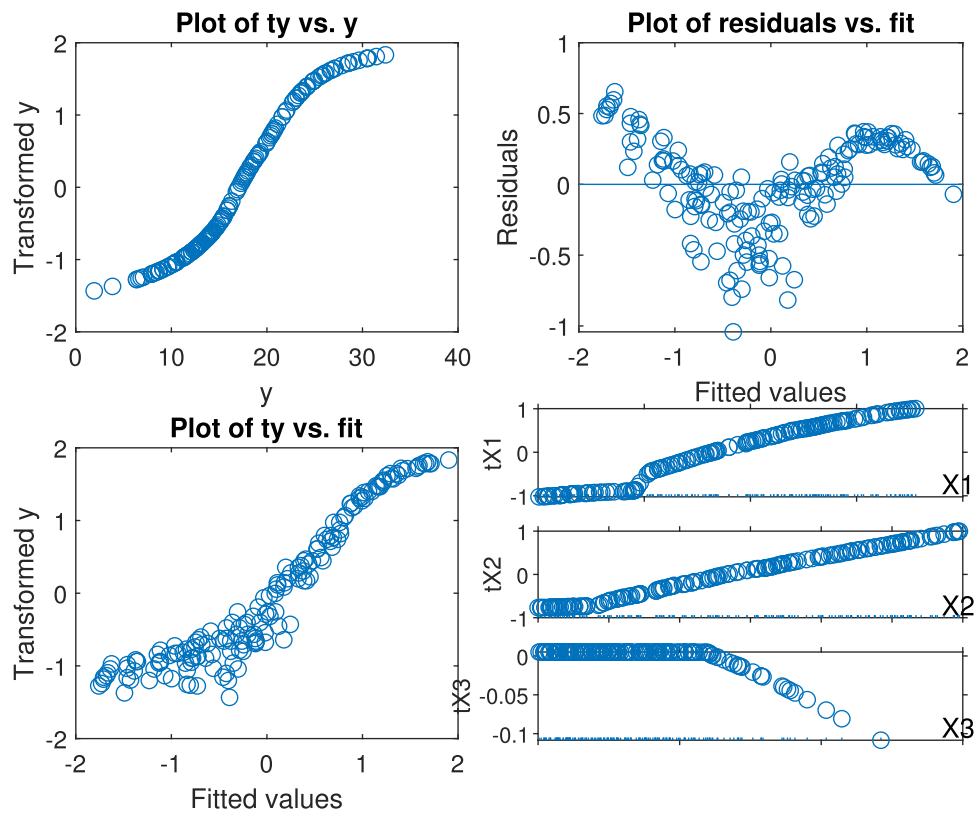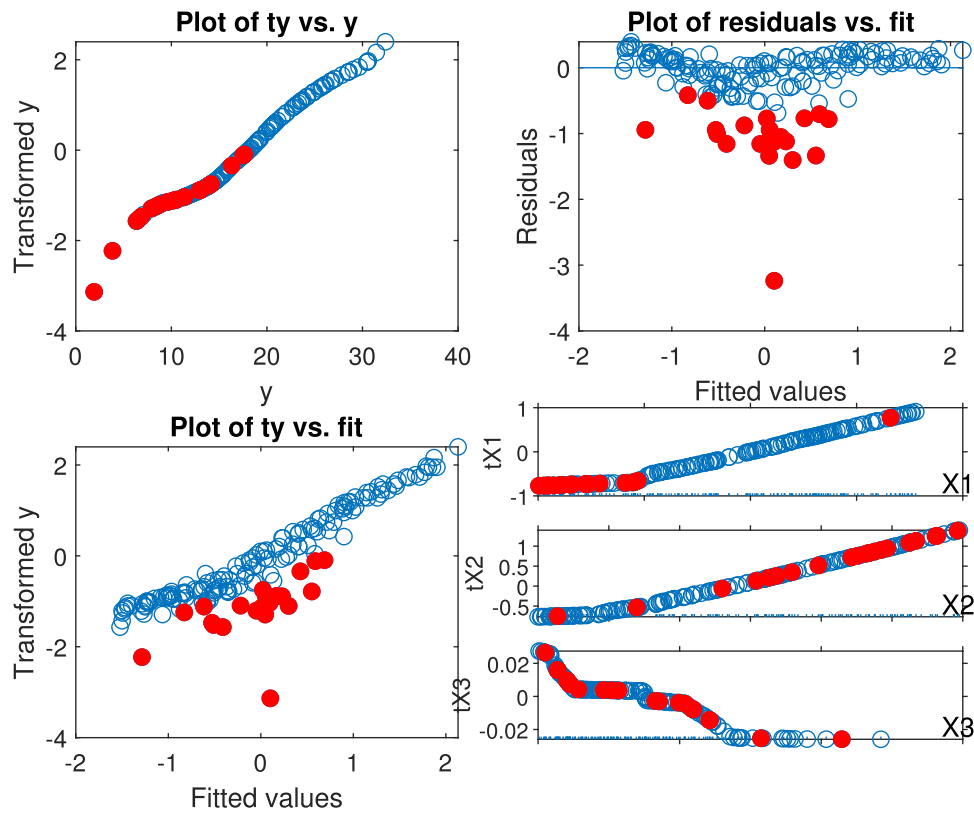

Figure 13. Marketing data. First-order model. Nonrobust analysis without options. Top left-hand panel, transformed *y* against *y*; top right-hand panel, residuals against fitted values; bottom left-hand panel transformed *y* against fitted values; bottom right-hand panel, transformations of the three explanatory variables.

AVAS is unstable unless the sample size exceeds 350. We used simulation to explore the properties of our method, comparing RAVAS with traditional AVAS, that is with no options. We found no numerical instability in either procedure for sample sizes from 200 to 10,000 and values of p from 5 to 20 with 10% additive outliers. The mean squared error of parameter
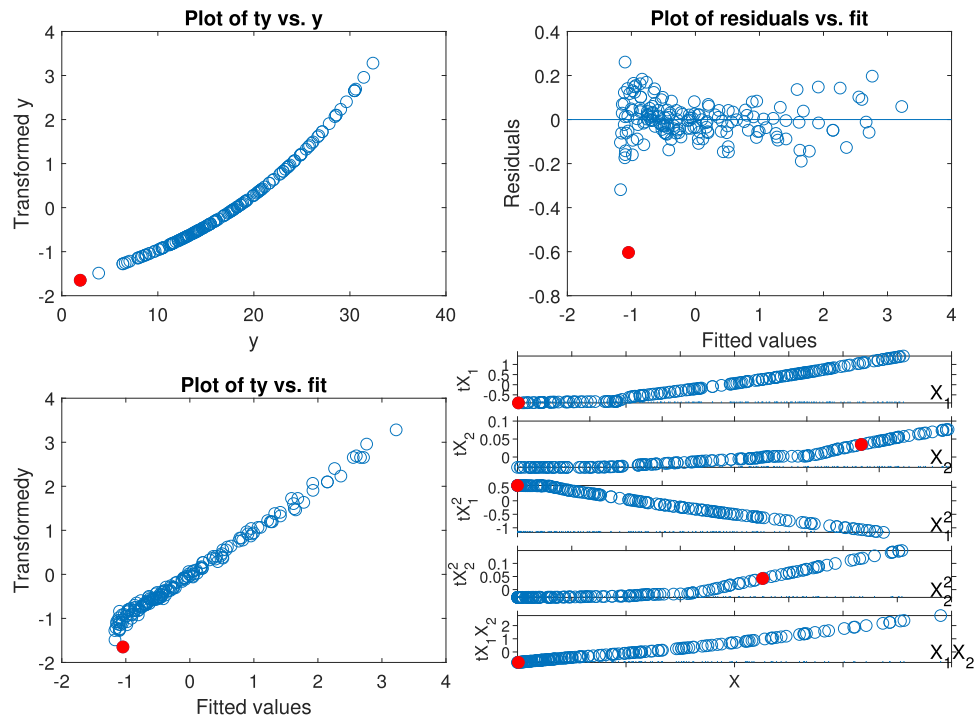
Figure 14. Marketing data. First-order model. RAVAS analysis with all options. Top left-hand panel, transformed *y* against *y*; top right-hand panel, residuals against fitted values; bottom left-hand panel transformed *y* against fitted values; bottom right-hand panel, transformations of the three explanatory variables. The 17 outliers shown by filled symbols.
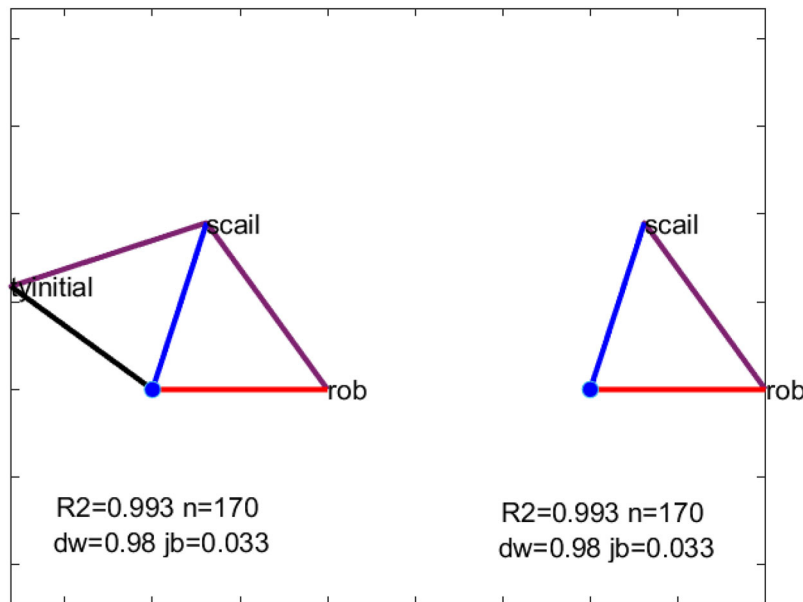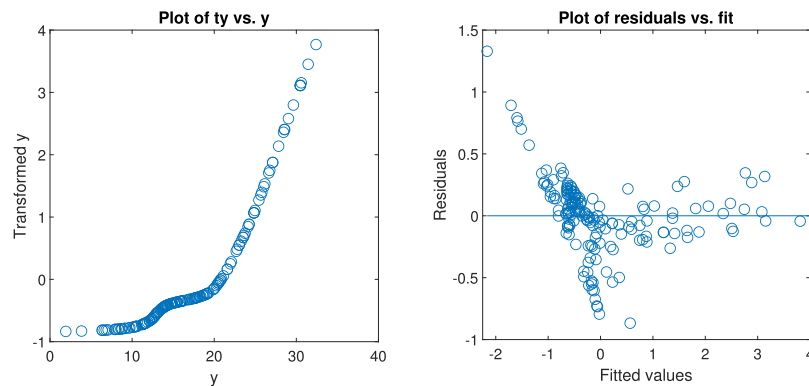

Figure 15. Marketing data. Quadratic model in x_1 and x_2. Augmented star plot with five potential options.

estimates from RAVAS remained virtually constant as the outliers became more remote, whereas that for AVAS increased steadily. The test for the detection of outliers using RAVAS tended to power one as the outliers became more remote. The average number of iterations for AVAS increased with sample size, and outlier remoteness, reaching the allowed upper limit of 20. The average number of iterations for RAVAS only rose

above three for small sample sizes and moderate shift contamination, the situation in which outliers cause significant bias in parameter estimates, but are hard to detect. We believe we have provided a robust version of AVAS in the spirit of the remark by Buja and Kass (1985) quoted in Section 1, that is a procedure that retains the computational simplicity of the original AVAS.

Figure 16. Marketing data, quadratic model. Analysis with options tyinitial, rob and scail. Top left-hand panel, transformed y against y; top right-hand panel, residuals against fitted values; bottom left-hand panel transformed y against fitted values; bottom right-hand panel, transformations of the five explanatory variables. The outlier is shown by a filled symbol.

Figure 17. Marketing data, quadratic model. Standard nonrobust analysis without options. Left-hand panel, transformed y against y; right-hand panel, residuals against fitted values.

Supplementary Materials

The first three sections of the supplementary material provide flow charts for the program we wrote to implement RAVAS. Section 2 is the initialisation of the fitting procedure including implementation of the four non-iterative options; Section 3 describes the iterative part of the algorithm and sets up the environment for the numerical variance stabilizing transformation; Section 4 provides the fitted values and residuals to which the trapezoidal integration approximation is applied; Section 5 provides links, inter alia, to the code used for the calculations and the plotting of graphs.

Acknowledgments

Disclosure Statement

Funding

References

Alimadad, A., and Salibian-Barrera, M. (2011), "An Outlier-Robust Fit for Generalized Additive Models with Applications to Disease Outbreak Detection," *Journal of the American Statistical Association*, 106, 719–731. [2]

Anscombe, F. J., and Tukey, J. W. (1963), "The Examination and Analysis of Residuals," *Technometrics*, 5, 141–160. [5]

Atkinson, A. C., and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, New York: Springer–Verlag. [5]

Atkinson, A. C., Riani, M., and Cerioli, A. (2010), "The Forward Search: Theory and Data Analysis ," (with Discussion), *Journal of the Korean Statistical Society*, 39, 117–134. DOI:10.1016/j.jkss.2010.02.007. [2,4]

Atkinson, A. C., Riani, M., and Corbellini, A. (2020), "The Analysis of Transformations for Profit-and-Loss Data," *Applied Statistics*, 69, 251–275. DOI:10.1111/rssc.12389. [4,7]

——— (2021), "The Box-Cox Transformation: Review and Extensions," *Statistical Science*, 36, 239–255. DOI:10.1214/20-STS778. [1]

Atkinson, A. C., Riani, M., Corbellini, A., Perrotta D., and Todorov, V. (2023), *Applied Robust Statistics through the Monitoring Approach*, Heidelberg: Springer Nature (in preparation). [4,5]

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D. (1972), *Statistical Inference under Order Restrictions*, Chichester: Wiley. [3]

Beaton, A. E., and Tukey, J. W. (1974), "The Fitting of Power Series, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," *Technometrics*, 16, 147–185. [4]

Boente, G., Martínez, A., and Salibián-Barrera, M. (2017), "Robust Estimators for Additive Models Using Backfitting," *Journal of Nonparametric Statistics*, 29, 744–767. [1,2]

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," (with Discussion), *Journal of the Royal Statistical Society*, Series B, 26, 211–252. [1]

Box, G. E. P., and Tidwell, P. W. (1962), "Transformations of the Independent Variables," *Technometrics*, 4, 531–550. [1]

Breiman, L. (1988), Comment on "Monotone Regression Splines in Action" (Ramsey, 1988), *Statistical Science*, 3, 442–445. [2,4]

Breiman, L., and Friedman, J. H. (1985), "Estimating Optimal Transformations for Multiple Regression and Transformation," (with discussion), *Journal of the American Statistical Association*, 80, 580–619. [1,3]

Buja, A., and Kass, R. E. (1985), "Comment on "Estimating Optimal Transformations for Multiple Regression and Transformation" by Breiman and Friedman," *Journal of the American Statistical Association*, 80, 602–607. [1,2,14]

Buja, A., Hastie, T., and Tibshirani, R. (1989), "Linear Smoothers and Additive Models," *Annals of Statistics*, 17, 453–510. [3]

Cantoni, E., and Ronchetti, E. (2001), "Robust Inference for Generalized Linear Models," *Journal of the American Statistical Association*, 96, 1022–1030. [2]

Cox, D. R., and Snell, E. J. (1981), *Applied Statistics: Principles and Examples*, London: Chapman and Hall. [5]

Durbin, J., and Watson, G. S. (1950), "Testing for Serial Correlation in Least Squares Regression: I," *Biometrika*, 37, 409–428. [7]

Foi, A. (2008), "Direct Optimization of Nonparametric Variance-Stabilizing Transformations," in *Meeting on Mathematical Statistics*. C.I.R.M. Marseille. [1]

Fowlkes, E. B., and Kettenring, J. R. (1985), "Comment on "Estimating Optimal Transformations for Multiple Regression and Transformation" by Breiman and Friedman," *Journal of the American Statistical Association*, 80, 607–613. [1]

Friedman, J., and Stuetzle, W. (1982), "Smoothing of Scatterplots," Technical Report, Department of Statistics, Stanford University, Technical Report ORION 003. [3]

Hampel, F. R. (1975), "Beyond Location Parameters: Robust Concepts and Methods," *Bulletin of the International Statistical Institute*, 46, 375–382. [4]

Harrell Jr, F. E. (2019), *Hmisc*, R package version 4.2-0. [2,12]

Hastie, T., and Tibshirani, R. (1986), "Generalized Additive Models," *Statistical Science*, 1, 297–318. [4]

——— (1988), Comment on "Monotone Regression Splines in Action" (Ramsey, 1988), *Statistical Science*, 3, 451–456. [3]

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman and Hall. [1,3]

Jarque, C. M., and Bera, A. K. (1987), "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 52, 163–172. [9]

Kassambara, A. (2019), *datarium: Data Bank for Statistical Analysis*, R package version 0.1.0. [12]

Ramsay, J. O. (1988), "Monotone Regression Splines in Action," *Statistical Science*, 3, 425–461. [2,4]

Riani, M., Atkinson, A. C., and Cerioli, A. (2009), "Finding an Unknown Number of Multivariate Outliers," *Journal of the Royal Statistical Society*, Series B, 71, 447–466. [4]

Riani, M., Atkinson, A. C., and Perrotta, D. (2014), "A Parametric Framework for the Comparison of Methods of Very Robust Regression," *Statistical Science*, 29, 128–143. [4]

Riani, M., Atkinson, A. C., and Corbellini, A. (2023), "Automatic Robust Box-Cox and Extended Yeo-Johnson Transformations in Regression," *Statistical Methods and Applications*, 32, 75–102. [4]

Rousseeuw, P. J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871–880. [4]

Rousseeuw, P. J., and Yohai, V. J. (1984), "Robust Regression by Means of S-estimators," in *Robust and Nonlinear Time Series Analysis: Lecture Notes in Statistics 26*, eds. J. Franke, W. Härdle, and R. D. Martin, pp. 256–272, New York: Springer Verlag. [2]

Schimek, M. G., and Turlach, B. A. (2006), "Additive and Generalized Additive Models," Technical report, Sonderforschungsbereich 373, Humboldt University, Berlin. [3]

Spector, P., Friedman, J., Tibshirani, R., Lumley, T., Garbett, S., and Baron, J. (2016), *acepack: ACE and AVAS for Selecting Multiple Regression Transformations*, R package version 1.4.1. [1,10]

Spiegel, E., Kneib1, T., and Otto-Sobotka, F. (2019), "Generalized Additive Models with Flexible Response Functions," *Statistics and Computing*, 29, 123–138. [2]

Tibshirani, R. (1988), "Estimating Transformations for Regression via Additivity and Variance Stabilization," *Journal of the American Statistical Association*, 83, 394–405. [1,3]

Torti, F., Corbellini, A., and Atkinson, A. C. (2021), "fsdaSAS: A Package for Robust Regression for Very Large Datasets Including the Batch Forward Search," *Stats*, 4, 327–347. [4]

Wang, D., and Murphy, M. (2005), "Identifying Nonlinear Relationships in Regression Ysing the ACE Algorithm," *Journal of Applied Statistics*, 32, 243–258. [10,12]

Wang, M., Lyu, B., and Yu, G. (2021), "ConvexVST: A Convex Optimization Approach to Variance-Stabilizing Transformation," in *Proceedings of the 38th International Conference on Machine Learning* (Vol. 139), pp. 10839–10848. [1]

Yohai, V. J., and Zamar, R. H. (1988), "High Breakdown-Point Estimates of Regression by Means of the Minimization of an Efficient Scale," *Journal of the American Statistical Association*, 83, 406–413. [2]