# Tests in the fan plot for robust, diagnostic transformations in regression

Anthony C. Atkinson [a,*], Marco Riani [b]

[a]*Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK*
[b]*Dipartimento di Economia, Università di Parma, Parma, Italy*

## Abstract

The fan plot of the score statistic for transformation during the forward search is a powerful tool for detecting masked outliers that indicate an incorrect transformation. We use simulation to investigate the distribution of this statistic during the search. Data structures where the normal approximation is, and is not, satisfactory are characterised. © 2002 Elsevier Science B.V. All rights reserved.

## 1. Introduction

Oftentimes, data need transforming before standard methods of analysis, such as regression, can be applied. The examples in this paper are time of survival and numbers of cycles to failure. Both are non-negative and, for example, the numbers of cycles range from 90 to 3636, so that a normal model with additive errors of constant variance is implausible. We use the Box and Cox parametric family of power transformations to find a suitable transformation of such data. Our interest is in the effect of outliers on estimation of this transformation and in the use of a robust, diagnostic method, the forward search, to find the transformation in the presence of multiple outliers.

Standard classical methods for finding outliers and influential observations start from a fit to all the data and then proceed backwards by the deletion of suspect observations. This procedure can be subject to masking, where several outliers hide the effect of each other, so that the data appear outlier-free. Robust methods, such as those described in Ref. [1], can overcome masking and reveal outliers. We use robustness, combined with the forward search, described in detail by Ref. [2], both to overcome masking and to reveal the importance of individual observations. The method starts from a small subset of the data and works forward by increasing the number of observations used in fitting. A useful diagnostic tool is to monitor the value of the score test for transformations during the search. Jumps in the plot indicate influential observations and may lead to the unmasking of multiple outliers.

In the forward plot, we judge the significance of the test statistic by reference to the normal distribution. The question posed in this paper is whether this

---

* Corresponding author. Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK. Tel.: +44-20-7955-7622; fax: +44-20-7955-7416.
*E-mail addresses:* a.c.atkinson@lse.ac.uk (A.C. Atkinson), mriani@unipr.it (M. Riani).

is an appropriate distribution. To find the answer, we use simulations to estimate the distribution of the statistic. The answer is that, in many examples, there is a negligible effect of the forward search on the statistic and that the normal approximation is a useful guide. The important corollary to this answer is that we are able to characterise those examples in which the normal approximation is poor.

In the next section, we introduce the test statistic for transformations. Section 3 describes the forward search. Examples of forward plots of the transformation statistic are in Section 4. Two ways in which the normal approximation may fail are characterised in Sections 5 and 6. The paper ends with a few general comments.

## 2. Transformations of the response in regression

Box and Cox [3] analyse the normalized power transformation

$$z(\lambda) = \begin{cases} (y^{\lambda} - 1)/(\lambda \dot{y}^{\lambda-1}) & \lambda \neq 0 \\ \dot{y} \log y & \lambda = 0, \end{cases} \tag{1}$$

where the geometric mean of the $n$ observations is written as $\dot{y} = \exp(\sum \log y_i/n)$. The hope is that, for some $\lambda$ which has to be estimated, the transformed observations will satisfy the linear regression model

$$z(\lambda) = x^T \beta + \epsilon, \tag{2}$$

where $x$ is $p \times 1$ and the errors $\epsilon$ are independently normally distributed with constant variance $\sigma^2$.

For inference about the transformation parameter $\lambda$, Box and Cox suggest the likelihood ratio test statistic. A disadvantage of the test is that a numerical maximisation is required to find the value of the maximum likelihood estimator $\hat{\lambda}$. For regression models, a computationally simpler alternative test of the hypothesis $\lambda = \lambda_0$ is the approximate score statistic derived by Taylor series expansion of Eq. (1) as

$$z(\lambda) \doteq z(\lambda_0) + (\lambda - \lambda_0)w(\lambda_0), \tag{3}$$

where

$$w(\lambda_0) = \left. \frac{\partial z(\lambda)}{\partial \lambda} \right|_{\lambda = \lambda_0}.$$

If the linearized response (3) is substituted in the regression model (2), the model becomes

$$z(\lambda_0) = x^T \beta - (\lambda - \lambda_0)w(\lambda_0) + \epsilon. \tag{4}$$

Because Eq. (4) is again a regression model with an extra variable $w(\lambda_0)$ derived from the transformation, the new variable is called the constructed variable for the transformation. If the true value of $\lambda$ is close to $\lambda_0$, the coefficient $(\lambda - \lambda_0)$ of the constructed variable will be small. The regression model (4) can be rewritten more conventionally by putting $\gamma = -(\lambda - \lambda_0)$ when

$$z(\lambda_0) = x^T \beta + \gamma w(\lambda_0) + \epsilon. \tag{5}$$

Small values of $\gamma$ then indicate that no transformation is necessary. The approximate score statistic $T_p(\lambda_0)$ for testing the transformation $\lambda = \lambda_0$ is just the $t$-statistic for the coefficient of regression on $w(\lambda_0)$ in Eq. (5). This can be calculated either directly from the multiple regression in Eq. (5), or from the formulae in Ref. [4, Chap. 6] in which multiple regression on $x$ is adjusted for the inclusion of the constructed variable. Such formulae give insight into the structure of the approximate score test but do not alter the numerical value of the statistic.

Similar ideas can be used for transformation of the explanatory variables [5]. Constructed variables for the joint transformation of the response and of explanatory variables, together with score tests, are given by Ref. [4, Section 8.4]. Whatever the formal statistical outcome of such procedures, the interpretability of the results is also of importance in finding a sensible transformation.

## 3. The forward search

### 3.1. General principles

Like most methods for outlier detection, our method divides the data into two parts: a larger "clean" part used for parameter estimation, and the outliers. The simplest division is into one potential outlier and the rest of the data, leading to single deletion diagnostics. Standard books on regression diagnostics, such as Refs. [4,6,7], include formulae for multiple deletion diagnostics, in which a small number, perhaps two or three, of potential outliers are

considered at once. However, there is a combinatorial explosion of the number of cases that have to be considered by such backwards working. We instead employ very robust methods that perform well even if almost half the observations are outliers.

We start our search with the resampling algorithm for least median of squares regression [8] in which the model is fitted to $p$ observations. In the forward search, larger subsamples of outlier free observations are found by starting from this subset and moving to larger subsets containing observations which have small residuals, and so are unlikely to be outliers. We use least squares to estimate the parameters from the selected subset of $m$ observations, obtaining vectors of estimates $\hat{\beta}(m)$ with $m$ running from $p$ to $n$. From these parameter estimates, we calculate a set of $n$ residuals. For each $m$, we plot the score statistic for transformations $T_p(\lambda)$, monitoring changes associated with the introduction of a particular observation into the subset $m$ used for fitting.

Since the forward search is described in detail in Ref. [2], only an outline is given here of the three steps.

### 3.2. Step 1: choice of the initial subset

If the model contains $p$ parameters, our forward search algorithm starts with the selection of a subset of $p$ units, which give an approximation to the least median of squares estimate. Observations in this subset are intended to be outlier free. If $n$ is moderate and $p \ll n$, the choice of the initial subset can be performed by exhaustive enumeration of all $\binom{n}{p}$ distinct $p$-tuples; otherwise, we evaluate the properties of some large number of $p$-tuples, often 1000. If $e_{i,S_\ell^p}$ is the least squares residual for unit $i$ given observations in $S_\ell^p$ are used in fitting, the initial subset $S_*^p$ is such that

$$e_{[\text{med}],S_*^p}^2 = \min_\ell \left[ e_{[\text{med}],S_\ell^p}^2 \right], \tag{6}$$

where $e_{[l],S_\ell^p}^2$ is the $l$th ordered squared residual among $e_{[i],S_\ell^p}^2$, $i = 1,\ldots,n$,

$$\text{med} = [(n + p + 1)/2], \tag{7}$$

and $[q]$ denotes the integer part of $q$. Criterion (6) provides a least median of squares method for regres-

sion models with independent errors [8,9]. The breakdown point of this estimator is asymptotically 50%. The forward search is not sensitive to the method used to select the initial subset: for example, least trimmed squares could be used in which the median in Eq. (6) is replaced by summation over a fraction of the observations.

### 3.3. Step 2: adding observations during the forward search

Given a subset $S_*^m$ of $m \geq p$ observations, the forward search selects the $m + 1$ units with the smallest squared least squares residuals, the units being chosen by ordering all squared residuals $e_{i,S_*^m}^2$, $i = 1,\ldots,n$.

In most moves from $m$ to $m + 1$, just one new unit joins the subset. However, sometimes, two or more units may join $S_*^m$ as one or more leave. Such an event is unusual, only occurring when the search includes one unit that belongs to a cluster of outliers. At the next step, the remaining outliers in the cluster seem less outlying and so several may be included at once. Of course, several other units then have to leave the subset. Step 2 of the forward search is repeated until all units are included in the subset.

### 3.4. Step 3: monitoring the search

There are three particular consequences of this procedure:

(1) *Stability.* In the absence of outliers and systematic departures from the model, the parameter estimates for each $m$ are unbiased estimators of the same quantity. So, both parameter estimates and residuals should remain approximately constant during the forward search.

(2) *Ordering.* If there are $k$ outliers and we start from a clean subset, the forward procedure will include these outliers towards the end of the search, usually in the last $k$ steps. Until then, residual plots and parameter estimates will remain approximately constant.

(3) *Transformations.* Outliers in one transformed scale may not be outliers in another scale. If the data are analysed using the wrong transformation, the $k$ outliers may enter the search well before the end. We therefore need to analyse the data using several values

of $\lambda$. We have given the name "fan plot" to the simultaneous forward plot of the score statistic $T_p(\lambda)$ for several values of $\lambda$.

## 4. Examples of transformation of the response

Our interest is in the distribution of the test statistic $T_p(\lambda)$ as the forward search progresses. To motivate this interest, we give two examples of fan plots. However, in this section, we focus our attention on the properties of the plot for a single value of $\lambda$. The distribution of the statistic is estimated by simulating 10,000 samples of data and performing one forward search on each from the best of 100 initial subsets. The examples, without simulation envelopes, and the fan plots are fully discussed in Chap. 4 of Ref. [2]. The extension to multivariate data is in Ref. [10].

### 4.1. Wool data

The first example is the wool data from Ref. [3] which are the number of cycles to failure of a worsted yarn under repeated loading. The results are from a single $3^3$ factorial experiment so that $n = 27$. The non-negative response ranges from 90 to 3636, so we would expect that a transformation would be appropriate. Box and Cox recommend the log transformation of $y$. This transformation is supported by a value of $-0.91$ for $T_p(0)$, the values of $-0.5$ and $0.5$ for $\lambda$ being rejected.

This analysis is uninformative about the contribution of individual observations to the transformation. Fig. 1 is the fan plot, that is the forward plot of the approximate score statistic $T_p(\lambda)$ for the five values of $\lambda$: $-1.0$, $-0.5$, $0$, $0.5$ and $1$. Each search is separate, so that the observations may, and do, enter in different orders in the five searches. The value of $T_p(-1)$ is at
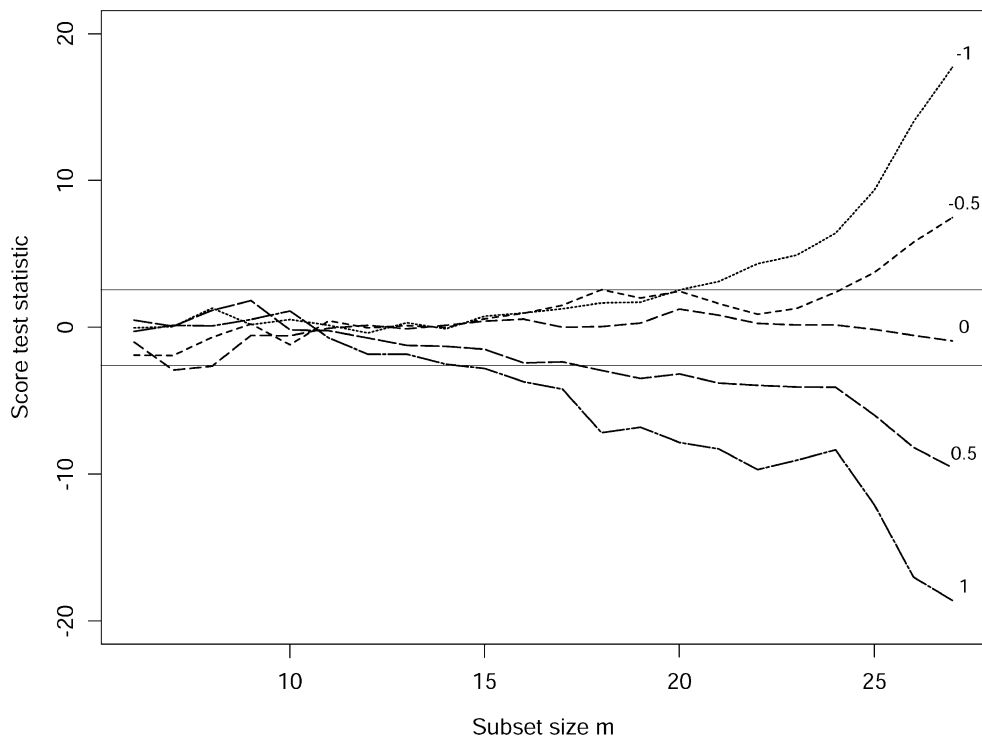


Fig. 1. Wool data: fan plot—forward plot of the score statistic $T_p(\lambda)$ for five values of $\lambda$ with theoretical 99% confidence band. Only the log transformation, $\lambda = 0$, is acceptable.

the top of the plot. The value of $T_p(0)$ lies within the central bands on the plot which are at $\pm 2.58$, containing 99% of the standard normal distribution. Provided the statistic has this null distribution, the indication is that the log transformation is acceptable. The other four curves move steadily outside this region as the search progresses, indicating the rejection of the other values of $\lambda$, the smoothness of the curves showing that this conclusion does not depend on one or a few outliers. The contrast with Fig. 4 is informative and is discussed in Section 4.3.

We now investigate the null distribution of $T_p(0)$. The score test is a *t*-test for regression on a constructed variable which is a function of the response. If this relationship between $y$ and $w$ is ignored, we would expect the score statistic to have a *t* distribution, apart from any effect of the ordering of observations due to the forward search. Fig. 2 shows the forward plot of $T_p(0)$ for the wool data during the forward search, together with the results of 1000 simulations when the data are generated with $\lambda = 0$

using the parameter estimates for this value of $\lambda$. The simulated 90, 95 and 99 percentage points of the distribution of the statistic show that the distribution of the statistic starts with longer tails than the normal but that, by half-way through this search, the distribution is close to the asymptotic standard normal distribution. The agreement is very good until the end of the search when there is a slight spreading of the distribution, recalling the bell of a trumpet. This slightly larger variance when $m = n$ is in line with the simulation results of Ref. [11].

### 4.2. Poison data

Our second example is the Poison data from Ref. [3]. The observations are time to death of animals in a $3 \times 4$ factorial experiment with four observations at each factor combination. Box and Cox suggest the reciprocal transformation ($\lambda = -1$), so that death rate, rather than survival time, has a simple structure. Fig. 3 shows the evolution of $T_p(-1)$ during the forward
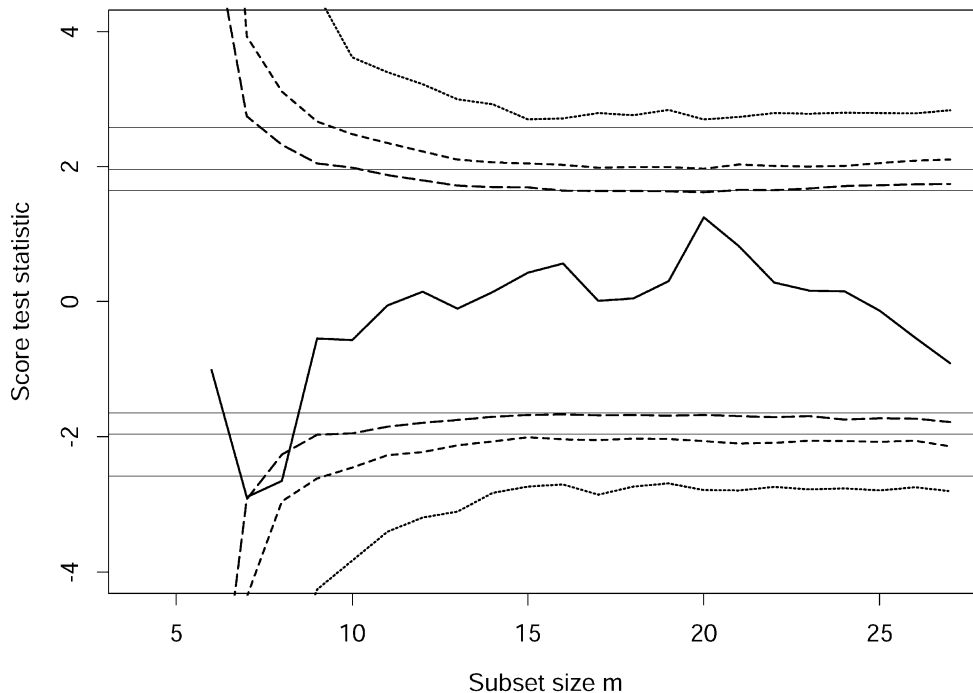


Fig. 2. Wool data: forward plot of $T_p(0)$ with 90%, 95% and 99% theoretical confidence bands and simulation envelopes using parameter estimates $\hat{\beta}(n)$ from the end of the search when $\lambda = 0$.
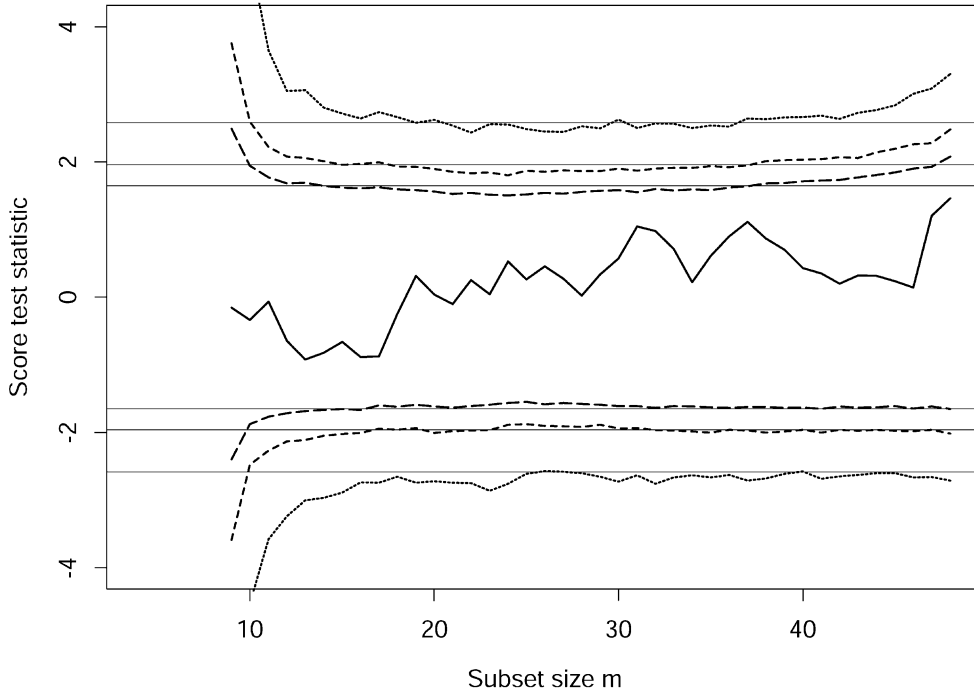
Fig. 3. Poison data: forward plot of $T_p(-1)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $\hat{\beta}(n)$.

search. The values are in the centre of the region and the transformation is supported by the plot. Again, we see good agreement between the simulated distribution and the normal distribution. Here, there is some slight evidence of a longer upper tail to the distribution towards the end of the search, although not enough to cause difficulty in interpreting the plot. We return to consideration of this slight skewness in Section 6.

### 4.3. Multiple modified Poison data: an example of masking

Our examples so far have shown that the plots can be well behaved in the absence of outliers. Atkinson and Riani [2] modified the Poison data to create four masked outliers which are not revealed by single deletion diagnostics and which indicate the incorrect transformation of $\lambda = 1/3$. Fig. 4 is the fan plot for these modified data, which is quite different from the fan plot for the wool data in Fig. 1. The curves are no longer smooth. That for $T_p(-1)$, reflecting the correct transformation for all but four observations, lies within the central boundary until the inclusion of the last four observations, which are the outliers, when the reciprocal transformation is overwhelmingly rejected. The plot clearly indicates that these four masked outliers are causing rejection of the transformation that is supported by the bulk of the data. The effect is similar, but less dramatic, for $T_p(-0.5)$. The score statistics for the other values do not have such simple shapes—they enter and leave the central boundaries as the outliers enter these searches before the end. A more detailed description of this plot starts on p. 106 of Ref. [2].

Since the overall analysis suggests $\lambda = 1/3$, we now look at the forward plot of $T_p(1/3)$ which is in Fig. 5. This also, of course, shows the effect of the four masked outliers, but in a different way. The outliers now enter towards the end of the forward search but are not the last four observations to enter. They cause the score statistic for this incorrect transformation to lie in the centre of the null distribution at the end of the search. The simulated values agree reasonably
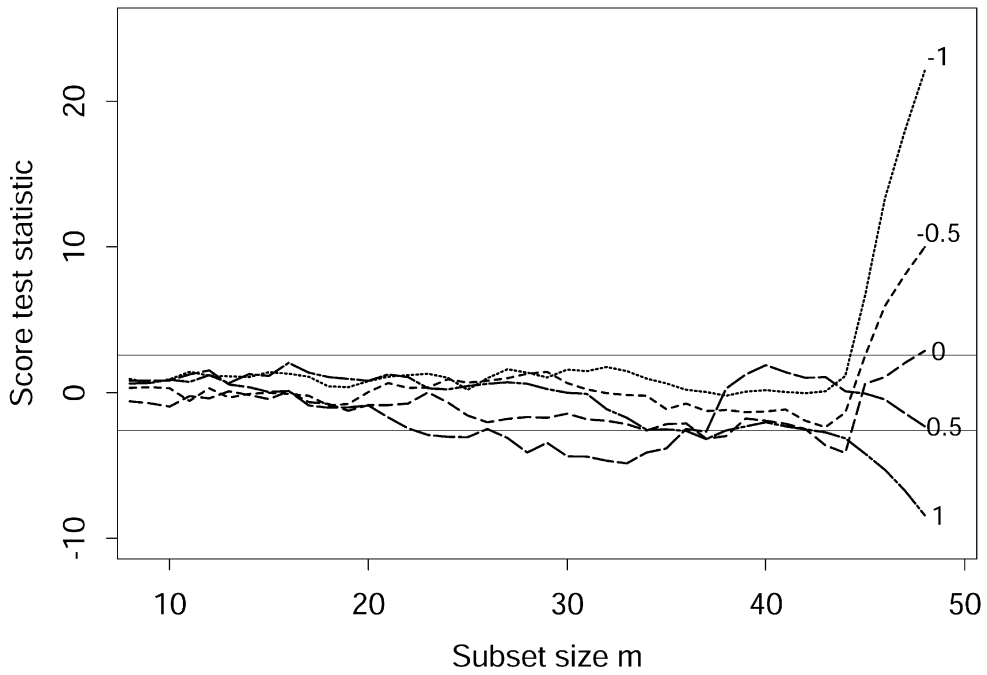
Fig. 4. Multiple modified Poison data: fan plot. The effect of the four outliers on the correct transformation for $\lambda = -1$ is evident.
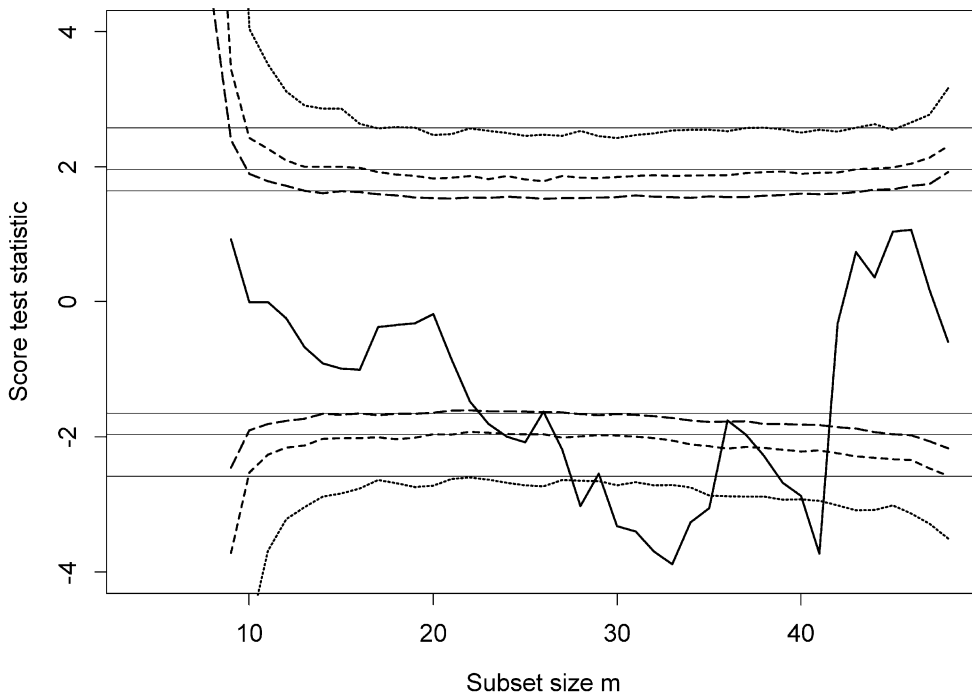


Fig. 5. Multiple modified Poison data: forward plot of $T_p(1/3)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $\hat{\beta}(n)$. The effect of the outliers in producing an erroneous transformation for all the data is evident.
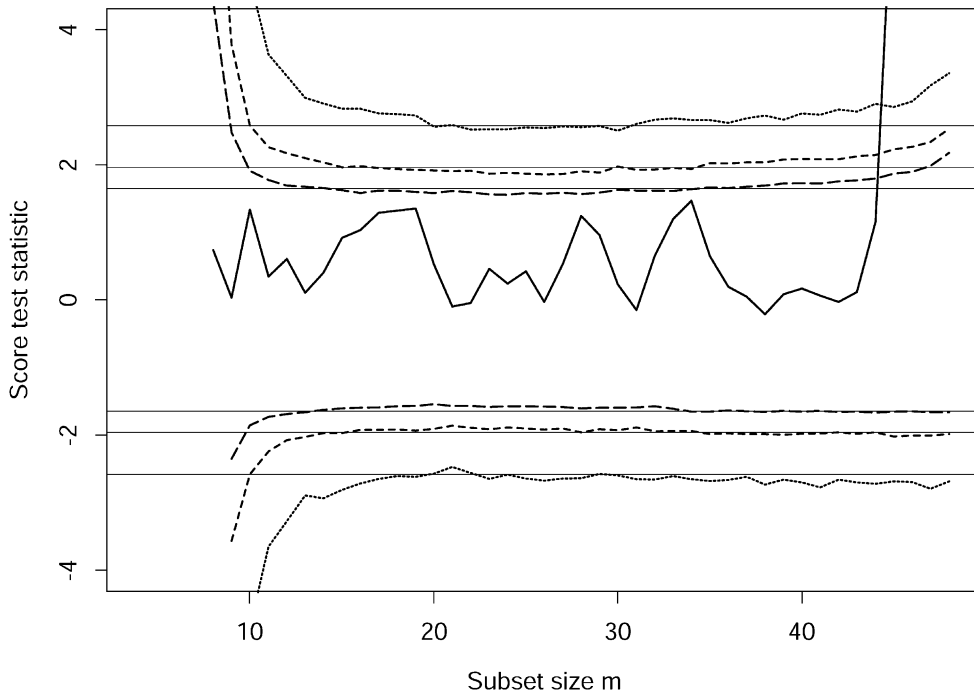
Fig. 6. Multiple modified Poison data: forward plot of $T_p(-1)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $\hat{\beta}(n-4)$ when $\lambda = -1$, that is after deletion of the four outliers.

with the normal quantiles, except for a slight downward bias at the end of the search together with a slight trumpet.

The effect of the four outliers is, as we have seen in the fan plot, overwhelming on $T_p(-1)$, given again in Fig. 6. The simulation envelopes shown in the figure, which agree well with the theoretical values, apart from a slight upward bias towards the end of the search, used the parameter estimates from $m = n - 4$, that is when all the outliers have been rejected.

## 5. Trumpets and constructed variables

In this section, we look at the relationship between the simulation envelopes and the parameter values used in the simulations. In envelopes of residuals in regression [4, p. 35] the linear model used for simulation does not matter and a sample of standard normal variables is used. What does matter is the hat matrix of the model fitted to the data, which affects the variances and covariances of the residuals.

However, in transformation, the parameters of the linear model can also have an effect. An obvious effect is that, for a power transformation to be considered, all observations must be positive. This effect is studied in the next section. Here, we study the rather more subtle effect of the dependence of the trumpet at the end of the search on the values of the parameters.

The effect of the trumpet was smallest in Fig. 2, for which the squared multiple correlation coefficient $R^2$ had the value 0.97 at the end of the search. For the other plots, $R^2$ was slightly less. To see whether $R^2$ is a useful way of characterising plots, Fig. 7 shows, for the wool data, simulation envelopes in which the linear model has the same constant and value of $s^2$ as those in the data, but in which the values of the remaining three parameters $b$ in the linear model have been divided by 10, that is, $b = \hat{\beta}(n)/10$. As a result, the average value of $R^2$ in the simulated data sets is reduced to 0.28. The effect on the simulation envelopes, compared to those in Fig. 2, is clear. Although symmetrical, the envelopes are now too large through-
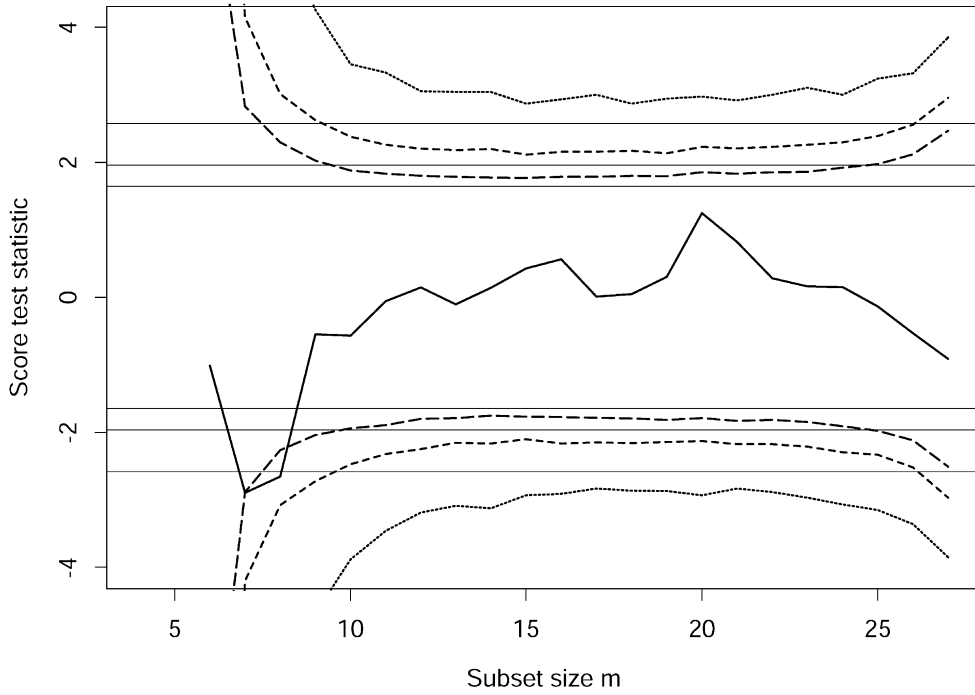
Fig. 7. Wool data: forward plot of $T_p(0)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $b = \hat{\beta}(n)/10$. There is now a trumpet towards the end of the envelopes.
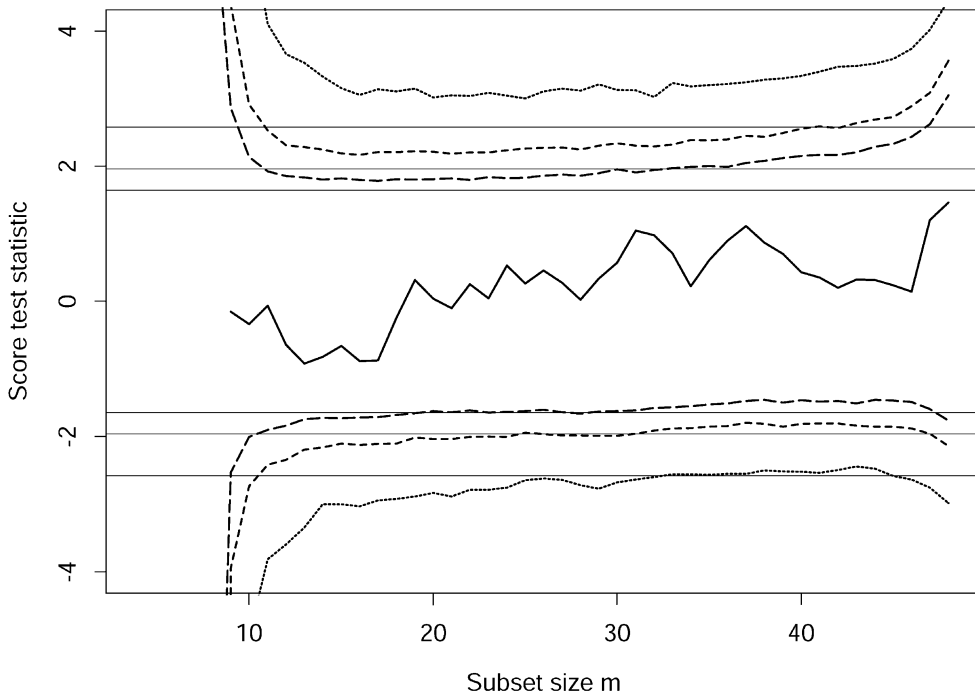


Fig. 8. Poison data: forward plot of $T_p(-1)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $b = \hat{\beta}(n)/10$. There is now an increase in both skewness and the trumpet effect.

out, especially towards the end of the search, where there is an appreciable trumpet.

A similar plot for the Poison data is given in Fig. 8. The effect of division by 10 is to reduce the value of $R^2$ from 0.85 to an average of 0.15. There is now a clear upper trumpet and the asymmetry of the simulation envelope has increased. Similar results are found for the stack loss data, but not presented here.

That a low value of $R^2$ accompanies wide simulation envelopes at the end of the search can be explained by considering the structure of the constructed variable plots, which are scatter plots of residual transformed response against the residual constructed variable. The score statistic $T_p(\lambda)$ is the $t$-test for interceptless regression in this plot. In the absence of evidence for a transformation, this plot often looks like a random scatter of points. However, in simple cases, the plots can have a near parabolic structure, even when there is no evidence for a transformation. Examples, some discussion and a Taylor series justification for the parabolic structure are given in Ref. [4, p. 192].

Panel (a) of Fig. 9 shows the forward plot of $T_p(-1)$ for the Poison data when only a constant is fitted. Although the transformation is correct when the factorial model is fitted, the inadequate model without effects seems to lead to the rejection of this value of $\lambda$. The other two panels are constructed variable plots. In both plots, the points lie on a nearly parabolic curve. Panel (b) of the plot is for the $m = n - 9$ pairs of residuals at that point in the forward search. It is nearly symmetrical, giving a value close to zero for the score test. In Panel (c), for all the data, there is asymmetry which will be reflected in the non-zero value of the score test. This figure shows both the structure of the plots and the way in which the ordering of the observations by the search causes extreme observations on the parabola to enter towards the end of the search.

If there are several explanatory variables and strong regression, the parabolic pattern disappears. However, if there is weak regression, giving a low
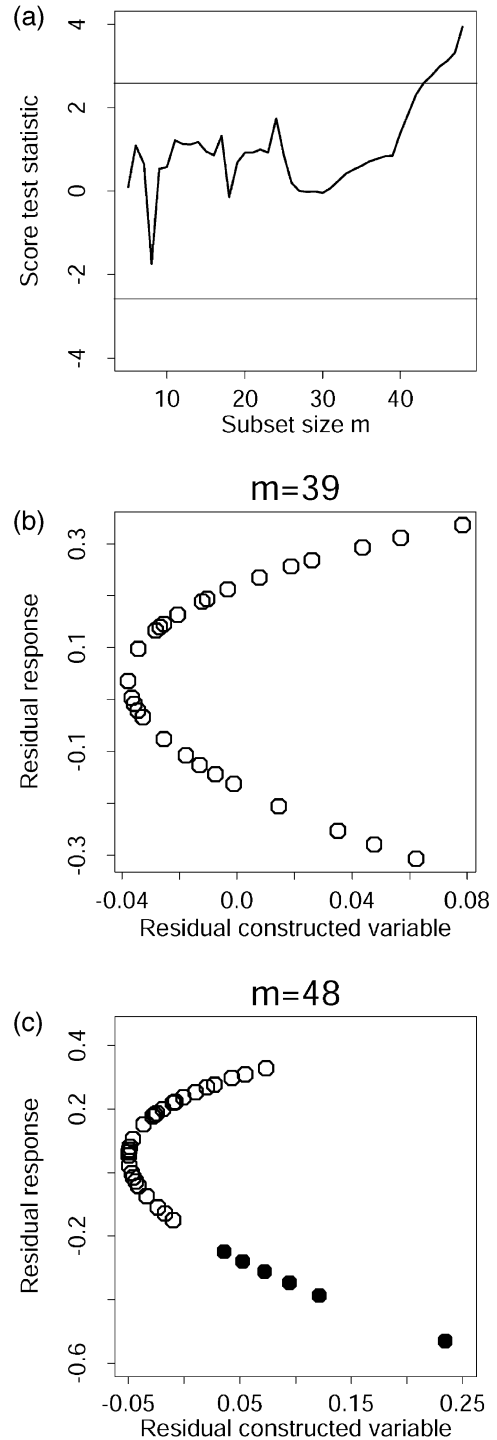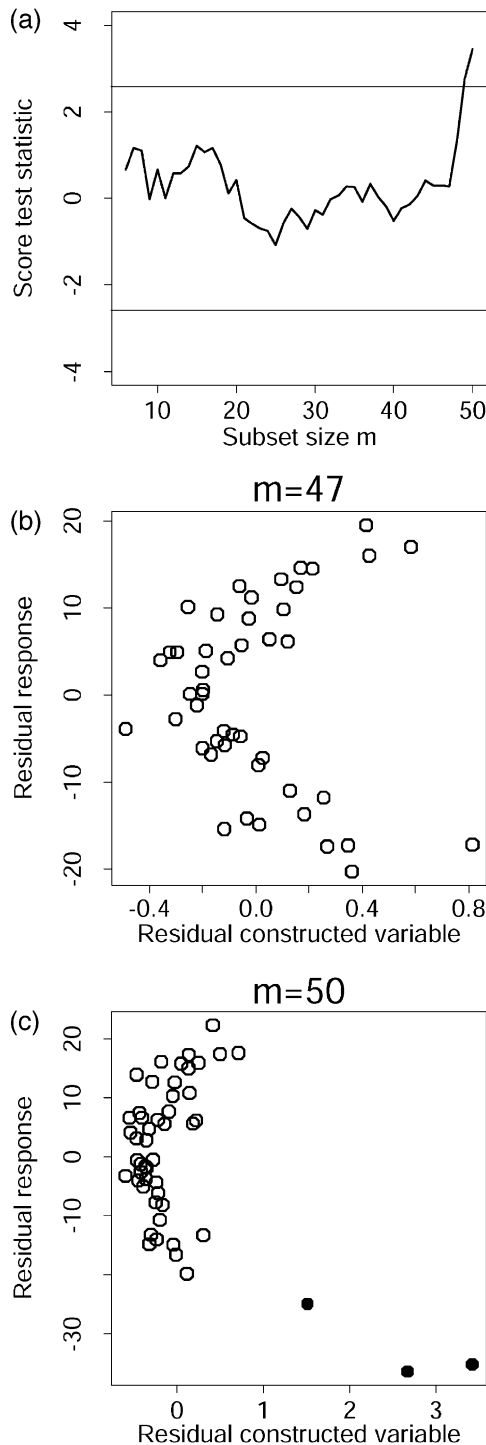


Fig. 9. Poison data with $\lambda = -1$ when only a constant is fitted: panel (a) forward plot of $T_p(-1)$ with theoretical 99% confidence band; panel (b) added variable plot for $m = n - 9$; panel (c) added variable plot for all the data: the filled symbols are the last observations to enter (some symbols overlap).

value of $R^2$, the effect is to jitter the parabola, the curves for groups of observations with a common fitted value being slightly displaced. An example is shown in the three panels of Fig. 10. The data were a simple random sample of 50 observations to which an intercept and three randomly generated explanatory variables were fitted. As a result, the value of $R^2$ is approximately zero.

Panel (a) shows a forward plot of $T_p(1)$ which increases sharply at the end, although the data were not simulated to need a transformation. This behaviour is close to that of the plot in Panel (a) of Fig. 9, suggesting that the evidence for rejection of the reciprocal transformation may not be as strong as it appeared. Panel (b) shows the constructed variable plot when three observations have been removed ($m = n - 3$). A symmetrical plot is obtained. The last panel for $m = n$ shows a blurred parabola which is asymmetric, the three observations causing the regression on the constructed variable being clearly visible. Overall, these plots show how, in the absence of regression structure in the data, spurious structure can occur in the constructed variable plots. Averaging over a number of simulations, many of which have this structure, gives rise to the trumpet in the simulation envelope observed in Figs. 7 and 8 in which the value of $R^2$ had been made artificially small. These results are a reminder that the evidence for a transformation depends on the structure of the linear model, if any, as well as on the distribution of errors.

## 6. Biases

Finally, we look at the biases in the distribution of the statistic that can be introduced by an inappropriate model.

Fig. 11 is the forward plot of $T_p(-1)$ for the multiple modified Poison data when the parameter estimates used in the simulation are those at the end of the search when the four outliers are included. As in Fig. 6, the observed value of the score statistic shoots up at the end of the plot. However, now, there is a

Fig. 10. Simulated data with no structure: panel (a) forward plot of $T_p(-1)$; panel (b) added variable plot for $m = n - 3$; panel (c) added variable plot for all the data. The filled symbols are the last observations to enter.
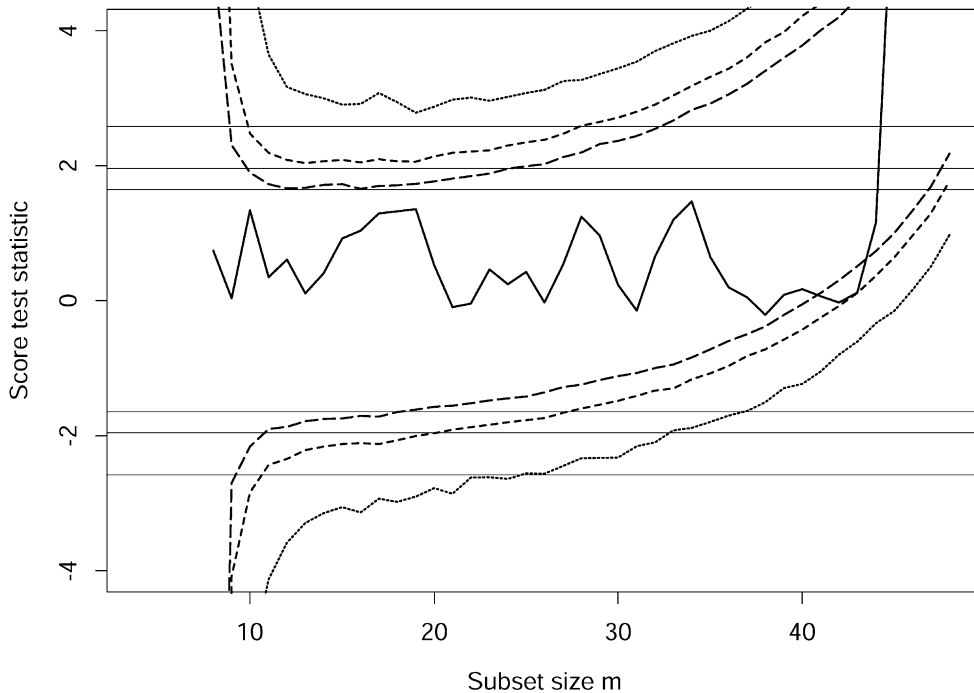
Fig. 11. Multiple modified Poison data: forward plot of $T_p(-1)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $\hat{\beta}(n)$ from $\lambda = -1$.

strong curve in the simulation envelope, which is far from having a zero mean, even though the simulations were performed using normal random variables and with $\lambda = -1$.

The explanation for this behaviour lies in the approximate nature of the Box–Cox procedure. For the transformation to be possible, the observations have to be positive and so have a truncated, rather than full, normal distribution. In our calculations, simulated values giving negative values of $y$ were rejected and resampled until a positive value was obtained. In generating Fig. 11, less than one-tenth of the simulated values could be used, so that sampling was very much skewed from a tail of the normal distribution. Because of the skewed shape of this conditional distribution, there is an apparent need for transformation, reflected in the positive values in the envelope towards the end of the search.

This explanation of skewness applies also to the lesser asymmetries seen in the envelopes in some of the other plots. To produce the 10,000 samples on which Fig. 11 is based, 113,266 negative simulated

values of $y$ were rejected. For other plots, the numbers were much more modest, around 900 for the two simulations of the unmodified Poison data in Figs. 3 and 8, and 757 for the multiple modified Poison data in Fig. 6 with parameter estimates from $m = n - 4$. A final plot is again of the multiple modified Poison data again using the parameter estimates from the last stage of the search, but now with $\lambda = -0.5$. With this value of $\lambda$ for simulation, there is an identically zero probability of generating a negative observation. As Fig. 12 shows, the simulation envelope is now well behaved, similar to that in Fig. 6 when $\lambda = -1$ but the parameter values used are for $m = n - 4$. In this new plot, the observed value of the statistic trends downward, suggesting ultimate rejection of $\lambda = -0.5$, a pattern which is destroyed by the presence of the outliers.

The comparison of the two sets of simulation envelopes in Figs. 11 and 12 confirms the importance of the truncated distribution of simulated values in skewing the distribution of the statistic. That there is an identically zero probability of generating negative values in the simulations for some values of $\lambda$ comes
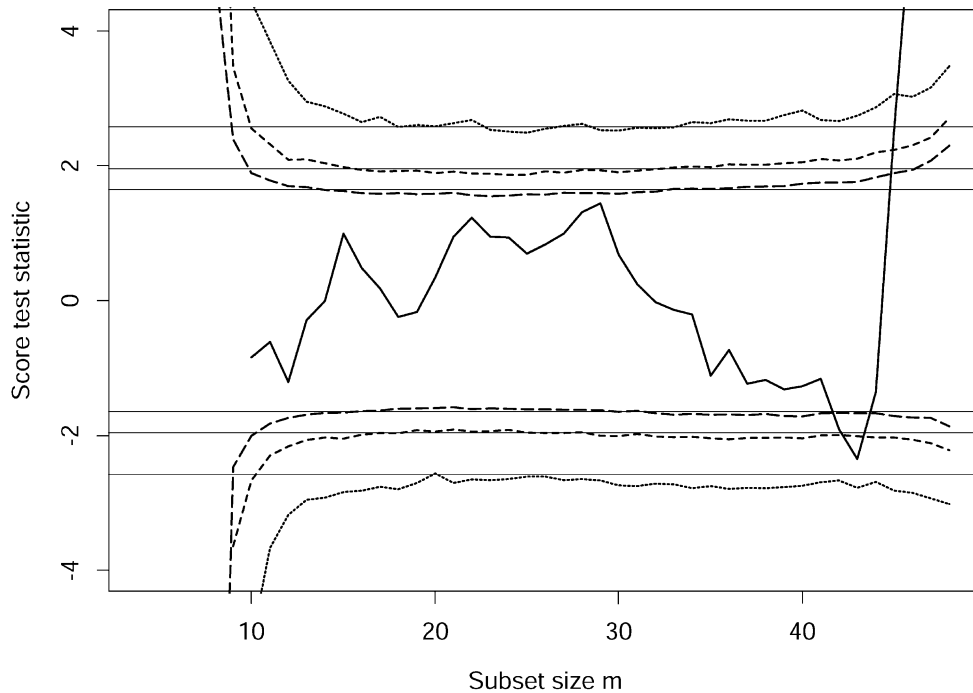
Fig. 12. Multiple modified Poison data: forward plot of $T_p(-0.5)$ with 90%, 95% and 99% simulation envelopes using parameter estimates $\hat{\beta}(n)$ for $\lambda = -0.5$. To be compared with Fig. 11.

from inverting the regression model (2) used for simulation. From the expression for the normalized transformation (Eq. (1)), the simulated values satisfy

$$y = [1 + \lambda \dot{y}^{\lambda-1}(x^T b + \epsilon)]^{1/\lambda}, \tag{8}$$

where $b$ is the parameter value used in the simulation. If, for example, $\lambda = 0.5$, $y$ is found by squaring the normal random variable on the right-hand side of Eq. (8), and so is never negative, even when the right-hand side is. However, there is also a source of slight bias here. When the data are transformed back to approximate normality, information on the sign of the square root is not available. Thus, if there is a non-negligible probability of negative values being generated before squaring, the distribution of $\sqrt{y}$ will be a truncated normal with some extra mass just above zero. This source of non-normality will result in a bias in the simulated envelope. Negative values in Eq. (8) are avoided for the special values $\lambda = 0$, $\pm 1/2$, $\pm 1/4$, $\pm 1/6, \ldots$ If $\lambda = 0$, the power in Eq. (8) is replaced by exponentiation so that problems of truncation do not

occur. We have already noted the symmetrical envelopes found for the wool data in Figs. 1 and 7.

A last characteristic of interest is the sign of the biases which, from Eq. (8), will change with the sign of $\lambda$. Our results show that negative values of $\lambda$ give positively skewed distributions, whereas those for positive $\lambda$ are negatively skewed. Only when $\lambda = 0$ is the distribution symmetrical.

## 7. Discussion

We have shown that, in many cases, the normal approximation to distribution of the score test holds well enough to provide a useful guide for inference. There is often no, or a negligible, effect of the forward search on the distribution of the statistic.

We have also characterised cases where the simulated distribution may not hold. One is when the value of $R^2$ is low, so that the fitted model is explaining only a small part of the variation in the data. The other is when there is a non-negligible probability that neg-

ative values of the response could be generated by simulation from the fitted model, which requires that the fitted values $\hat{z}$ should be such that $\hat{z}/s > 3$, or so, where $s$ is the residual mean square estimate of the error standard deviation. The transformation model may be inappropriate for data for which this inequality does not hold. Both of these cases are easily identified. In their absence, simulation is not necessary and the standard normal distribution provides a good guide to the behaviour of the forward plot of the score statistic.

A final point is that we have studied the properties of plots for a single value of $\lambda$. In the analysis of data, it is important to use several values of $\lambda$ if masking is to be detected. The fan plot is one way of achieving this.

An S-Plus library for the forward search in regression, including fan plots, is available at our web site http://www.stat.econ.unipr.it/riani/ar.

## References

[1] A. Singh, J.M. Nocerino, Robust procedures for the identification of multiple outliers, in: J. Einert (Ed.), Handbook of Environmental Chemistry Vol. 2: Statistical Methods: Part G. Chemometrics Applications, Springer-Verlag, New York, 1995, pp. 229–277.

[2] A.C. Atkinson, M. Riani, Robust Diagnostic Regression Analysis, Springer, New York, 2000.

[3] G.E.P. Box, D.R. Cox, An analysis of transformations (with discussion), Journal of the Royal Statistical Society, Series B 26 (1964) 211–246.

[4] A.C. Atkinson, Plots, Transformations, and Regression, Oxford Univ. Press, Oxford, 1985.

[5] G.E.P. Box, P.W. Tidwell, Transformations of the independent variables, Technometrics 4 (1962) 531–550.

[6] R.D. Cook, S. Weisberg, Residuals and Influence in Regression, Chapman & Hall, London, 1982.

[7] S. Chatterjee, A.S. Hadi, Sensitivity Analysis in Linear Regression, Wiley, New York, 1988.

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

[9] D.M. Hawkins, The accuracy of elemental set approximations for regression, Journal of the American Statistical Association 88 (1993) 580–589.

[10] M. Riani, A.C. Atkinson, Robust diagnostic data analysis: transformations in regression (with discussion), Technometrics 42 (2000) 384–394.

[11] A.C. Atkinson, A.J. Lawrance, A comparison of asymptotically equivalent tests of regression transformation, Biometrika 76 (1989) 223–229.