Taylor & Francis
Taylor & Francis Group

ROBUST TRANSFORMATIONS IN UNIVARIATE AND MULTIVARIATE TIME SERIES

Marco Riani

*Division of Statistics and Computing, Department of Economics,
University of Parma, Parma, Italy*

□   *It is well known that transformation of the response may improve the homogeneity and the approximate normality of the errors. Unfortunately, the estimated transformation and related test statistic may be sensitive to the presence of one, or several, atypical observations. In addition, it is important to remark that outliers in one transformed scale may not be atypical in another scale. Therefore, it is important to choose a transformation which does not depend on the presence of particular observations. In this article we suggest an efficient procedure based on a robust score test statistic which quantifies the effect of each observation on the choice of the transformation.*

1. INTRODUCTION

The analysis of data is often improved by using a transformation of the response rather than the original response itself. However, finding a suitable transformation can be strongly affected by the influence of a few individual observations. These difficulties are enhanced in the case of time series models since any anomalies are with respect to the specific components of the model. Atkinson and Shephard (1996) develop single deletion diagnostics for the effect of individual observations on the estimated transformation in time series. As we will see, these methods are prone to masking and swamping with a cluster of outliers, but more generally are unable to identify the effect of subsets of observations on the choice of the transformation parameters. In this article, we show how

Address correspondence to Marco Riani, Division of Statistics and Computing, Department of Economics, University of Parma, Via Kennedy 6, Parma I-43100, Italy; E-mail: mriani@unipr.it

the forward search approach (Atkinson and Riani, 2000; Atkinson et al., 2004) can be extended to robustly estimate the Box–Cox transformation parameters and improve normality of univariate and multivariate time series data. The emphasis in this article is on the state space form, so that the outlined procedure can be easily applied both to multivariate autoregressive integrated moving average (ARIMA) and structural time series models.

The structure of the article is as follows. In Section 2 we provide a unified framework for multivariate transformations both for independent and correlated observations, and discuss the advantages and the drawbacks of the score test applied to multivariate time series. In Section 3 we describe our procedure for finding in a robust way the transformation parameters. In Section 4 we apply the suggested procedure to real time series and compare our results with those previously obtained in the literature. In Section 5 we investigate the forward distribution of the score test statistic for transformations in time series. Section 6 contains concluding remarks and directions for future research.

2. MULTIVARIATE TRANSFORMATIONS TO NORMALITY

In this section we initially assume that the observations are independent. Later we extend the results to multivariate time series. Let $(y_1, \ldots, y_T)$ be a random sample of T observations from a multivariate distribution of dimension N. Let Y be a $T \times N$ matrix of observations with tth row y_t, the jth element of which is y_{tj}. In the extension of the Box and Cox (1964) family to multivariate responses there is a vector λ of N transformation parameters λ_j, one for each of the N responses. The normalized transformation of y_{tj} is given by

$$
\begin{aligned}
z_{tj}(\lambda_j) &= \left(y_{tj}^{\lambda_j} - 1\right)/\lambda_j \dot{y}_j^{\lambda_j - 1} \quad (\lambda \neq 0) \\
&= \dot{y}_j \log y_{ij} \qquad\qquad (\lambda = 0),
\end{aligned}
\tag{1}
$$

where $\dot{y}_j$ is the geometric mean of the jth response. The value $\lambda_j = 1$ $(j = 1, \ldots, N)$ corresponds to no transformation of any of the responses. Similarly, $\lambda = 0$ corresponds to the log transformation for each variable and so on.

We assume a multivariate linear regression model of the form

$$
Z(\lambda) = (X_1 \beta_1, \ldots, X_N \beta_N) + \Xi,
\tag{2}
$$

where $Z(\lambda)$ is a $T \times N$ matrix of normalized responses whose tjth generic element $z_{tj}(\lambda_j)$ is defined in Eq. (1). The X_j, $j = 1, \ldots, N$ are $T \times p$ design matrices not necessarily equal. The β_j are unknown vectors of parameters and Ξ is a $T \times N$ random matrix whose rows are i.i.d.

If the transformed observations are normally distributed with mean μ_t for the tth observation and covariance matrix Σ, twice the profile loglikelihood of the observations is given by

$$2L_{\max}(\lambda) = \text{const} - T \log \left| \widehat{\Sigma}(\lambda) \right|$$
$$- \sum_{t=1}^{T} \{z_t(\lambda) - \hat{\mu}_t(\lambda)\}' \widehat{\Sigma}^{-1}(\lambda) \{z_t(\lambda) - \hat{\mu}_t(\lambda)\} \qquad (3)$$

$$= \text{const} - T \log \left| \widehat{\Sigma}(\lambda) \right| - \sum_{t=1}^{T} e_t(\lambda)' \widehat{\Sigma}(\lambda)^{-1} e_t(\lambda), \qquad (4)$$

where $z_t(\lambda) = (z_{t1}(\lambda_1), \ldots, z_{tN}(\lambda_N))^T$ denotes the tth row of matrix $Z(\lambda)$ and $e_t(\lambda)$ the $N \times 1$ vector of residuals for observation t for the same value of λ. The calculation of $\hat{\mu}_i(\lambda)$ and $\widehat{\Sigma}(\lambda)$ is simplified when the matrix of explanatory variables X is the same for all responses. As a result, the least squares estimates can be found by independent regression on each response, yielding the $p \times N$ matrix of parameter estimates $B(\lambda) = (\hat{\beta}_1, \ldots, \hat{\beta}_N) = (X^T X)^{-1} X^T Z(\lambda)$. In other words, if the explanatory variables are the same for all responses, the maximum likelihood estimators of β_j can be obtained regressing the jth column of $Z(\lambda)$ on X ($j = 1, \ldots, N$) (see for example Hamilton, 1994, p. 318). On the other hand, when the X_j are different we have a so called SUR type model and maximum likelihood estimation requires iteration (Zellner, 1962).

Once the parameter estimates $\hat{\mu}_t(\lambda)$ and $\widehat{\Sigma}(\lambda)$ are found, where

$$T\widehat{\Sigma}(\lambda) = \sum_{t=1}^{T} e_t(\lambda) e_t'(\lambda), \qquad (5)$$

we can substitute this estimate in (4). The profile loglikelihood reduces to

$$2L_{\max}(\lambda) = \text{const}' - T \log \left| \widehat{\Sigma}(\lambda) \right|. \qquad (6)$$

So, to test the hypothesis $\lambda = \lambda_0$, the likelihood ratio test statistic

$$T_{LR} = T \log \left\{ \left| \widehat{\Sigma}(\lambda_0) \right| / \left| \widehat{\Sigma}(\hat{\lambda}) \right| \right\} \qquad (7)$$

is compared with the χ^2 distribution on N degrees of freedom. In (7) $\hat{\lambda}$ is the vector of N parameter estimates maximizing (4), which is found by numerical search. Replacement of $\widehat{\Sigma}(\lambda)$ in (7) by the unbiased estimator $\widehat{\Sigma}_u(\lambda)$ results in the multiplication of each determinant by a factor which cancels, leaving the value of the statistic unchanged.

If the observations are not i.i.d., in order to take into account both the presence of underlying components (trend, seasonal, etc.) and

the autocorrelation structure of the time series, we have to put the model in the state space form. The Kalman filter provides a unifying tool for state-space model likelihood evaluation and prediction. Given that both ARIMA and structural time series models can be put into the state space formulation and can be estimated using the Kalman filter, the description which follows provides a unified framework for the analysis of transformations in multivariate time series. The extension of Eq. (2) to multivariate time series is the state space form which follows

$$z_t(\lambda) = Z_t \alpha_t + X_t \beta_t + G_t \epsilon_t \tag{8}$$

$$\alpha_{t+1} = T_t \alpha_t + W_t \beta_t + H_t \epsilon_t, \tag{9}$$

where $\epsilon_t \sim WN(0, \sigma^2 I)$. Generally, the system matrices Z_t, G_t, T_t, H_t, X_t, and W_t are functionally related to a vector of hyperparameters θ. The measurement Eq. (8) is similar to the multivariate regression model (2) except that now the parameters evolve according to the transition Eq. (9). Vector β_t in Eqs. (8) and (9) may contain random, diffuse effects (i.e., trend and seasonal) and/or fixed effects (i.e., explanatory variables). If it is equal to zero and $\alpha_1 \sim N(a_{1|0}, \sigma^2 P_{1|0})$ with $a_{1|0}$, $P_{1|0}$ and σ^2 known, the standard Kalman filter provides a recursive algorithm for computing the minimum mean squared error estimator of α_t conditional on $z_1(\lambda), \ldots, z_{t-1}(\lambda)$, i.e.,

$$a_{t|t-1} = E(\alpha_t \mid z_1(\lambda), \ldots, z_{t-1}(\lambda))$$

and its mean squared error (*MSE*)

$$MSE(a_{t|t-1}) = E[(a_{t|t-1} - \alpha_t)(a_{t|t-1} - \alpha_t)' \mid z_1(\lambda), \ldots, z_{t-1}(\lambda)] = P_{t|t-1}.$$

More specifically, the Kalman filter is the set of recursions

$$v_t(\lambda) = z_t(\lambda) - Z_t a_{t|t-1} \qquad F_t(\lambda) = Z_t P_{t|t-1} Z_t' + G_t G_t'$$

$$q_t = q_{t-1} + v_t'(\lambda) F_t^{-1}(\lambda) v_t(\lambda) \qquad K_t = (T_t P_{t|t-1} Z_t' + H_t G_t') F_t^{-1}(\lambda)$$

$$a_{t+1|t} = T_t a_{t|t-1} + K_t v_t(\lambda) \qquad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t' - K_t F_t(\lambda) K_t', \tag{10}$$

with $q_0 = 0$. The filter innovations (one step ahead prediction errors) are indicated by $v_t(\lambda)$ and their variance by $\sigma^2 F_t(\lambda) = \text{var}(v_t(\lambda)) = \text{var}\{z_t(\lambda) - E(z_t(\lambda)|z_1(\lambda), \ldots, z_{t-1}(\lambda))\}$. These two quantities form the necessary ingredients for the computation of the likelihood. The extension of Eq. (4) to multivariate time series is given by

$$2L_{\max}(\lambda, \theta) = \text{const} - NT \ln \sigma^2 - \sum_{t=1}^{T} \ln |F_t(\lambda)| - \sigma^{-2} q_T. \tag{11}$$

The maximum likelihood estimate of σ^2 in equation (11) is

$$\hat{\sigma}^2 = \frac{q_T}{NT} = \frac{\sum_{t=1}^{T} v_t(\lambda)'F_t^{-1}(\lambda)v_t(\lambda)}{NT}.$$

If $\hat{\sigma}^2$ is concentrated out of the likelihood, the profile maximized loglikelihood can be written as

$$2L_{\max}(\lambda, \theta) = \text{const}' - NT \ln \frac{\sum_{t=1}^{T} v_t(\lambda)'F_t^{-1}(\lambda)v_t(\lambda)}{NT} - \sum_{t=1}^{T} \ln |F_t(\lambda)|. \quad (12)$$

As in model (2), different values of λ can be compared by calculation of the profile loglikelihood (12). Note that while for a random sample the data enter the likelihood solely through the residual sum of squares of $e_t(\lambda)$ and their covariance matrix, in multivariate time series, the corresponding role of these quantities is played by $v_t(\lambda)$ and $F_t(\lambda)$.

The likelihood ratio to test $H_0 : \lambda = \lambda_0$ becomes

$$T_{LR} = NT \ln \frac{\sum_{t=1}^{T} v_t(\lambda_0)'F_t^{-1}(\lambda_0)v_t(\lambda_0)}{\sum_{t=1}^{T} v_t(\hat{\lambda})'F_t^{-1}(\hat{\lambda})v_t(\hat{\lambda})} + \sum_{t=1}^{T} \ln\{|F_t(\lambda_0)|/|F_t(\hat{\lambda})|\}. \quad (13)$$

A disadvantage of this test is that a numerical maximization is required to find the value of $\hat{\lambda}$. Given that for each univariate time series we already have to estimate iteratively the parameters, the computational burden added to estimate the transformation parameters is non trivial. A computationally simpler alternative is to use a score test derived by Taylor series expansion of Eq. (1). This linearization leads to the $T \times N$ values of the N constructed variables

$$w_{tj}(\lambda_0) = \left. \frac{\partial z_{tj}(\lambda)}{\partial \lambda_j} \right|_{\lambda = \lambda_0}$$

$$= y_{tj}^{\lambda_{0j}} \log y_{tj} / \left(\lambda_{0j} \dot{y}_j^{\lambda_{0j}-1} \right) - z_{tj}(\lambda_0)(1/\lambda_{0j} + \log \dot{y}_j), \quad (14)$$

in which the jth response is differentiated with respect to λ_j (Atkinson and Riani, 2000, Chapter 4).

The combination of (14) and the state space formulation of Eq. (8) yields the model

$$z_t(\lambda_0) = Z_t\alpha_t + X_t\beta_t + w_t'(\lambda_0)\gamma + G_t\epsilon_t \quad (15)$$

$$\alpha_{t+1} = T_t\alpha_t + W_t\beta_t + H_t\epsilon_t, \quad (16)$$

where $w_t = (w_{t1}(\lambda_0), \ldots, w_{tN}(\lambda_0))'$ and $\gamma = (\gamma_1, \ldots, \gamma_N)'$.

Equations (15) and (16) are again a new state space formulation with N extra variables $w_j(\lambda_0)$ derived from the transformation. The new variables are called *constructed variables* for transformation. Using the approximate score test to determine whether λ_0 is the correct transformation of the N responses is equivalent to testing that the N parameters $\gamma_j, j = 1, \ldots, N$ are all zero.[1] The approximate score statistic for testing the transformation, $T_{SC}(\lambda_0)$, is the F statistic in the system which includes the additional variables $w_j(\lambda_0)$ in (15). Of course, if $N = 1$ the score test is nothing but the t test statistic associated with the variable $w(\lambda_0)$. Vector w_t can be incorporated in matrix X_t to give

$$z_t(\lambda_0) = Z_t\alpha_t + X_t^*\beta_t^* + G_t\epsilon_t \tag{18}$$

$$\alpha_{t+1} = T_t\alpha_t + W_t^*\beta_t^* + H_t\epsilon_t. \tag{19}$$

Note that it is not restrictive that the parameter vector β_t^* appears in both the measurement and transition equations, because the matrices X^* and W_t^* can be interpreted as selection matrices. Some specific components of β^* (i.e., trend and seasonal effects) are associated only with the initial conditions α_0 and do not manifest themselves in $z_t(\lambda_0)$.

Up to now we have assumed that the initial conditions were known. With nonstationary elements in order to avoid rounding errors we have used the so called augmented diffuse Kalman filter (see De Jong and Chu Chun Lin, 1994a,b). Basically, this solution requires augmenting the vector recursions of the Kalman filter to matrix recursions and adding an extra matrix recursion. A transition to the usual Kalman filter after some initial point $t = d$ is optional (see also Koopman and Durbin, 2003).

It is important to notice that even if originally there were no explanatory variables, Eq. (18) due to the presence of the constructed variables becomes a time series analogue of the seemingly unrelated regression equation (SURE) model. This model is known in the time

---

[1]In order to better clarify this statement, let us consider a simple example. If we replace the model

$$z_t(\lambda) = \beta_0 + \beta_1 x_t + \epsilon_t$$

with the linear approximation

$$z_t(\lambda_0) + (\lambda - \lambda_0)w_t(\lambda_0) \approx \beta_0 + \beta_1 x_t + \epsilon_t,$$

then

$$z_t(\lambda_0) = \beta_0 + \beta_1 x_t - (\lambda - \lambda_0)w_t(\lambda_0) + \epsilon_t$$
$$= \beta_0 + \beta_1 x_t + \gamma w_t(\lambda_0) + \epsilon_t, \tag{17}$$

where $\gamma = -(\lambda - \lambda_0)$. The t test for $\gamma = 0$ ($T_{SC}(\lambda_0)$) in (17) is the test of the hypothesis $\lambda = \lambda_0$. Because $T_{SC}(\lambda_0)$ is the t test for regression on $-w(\lambda_0)$, large positive values of the statistic mean that λ_0 is too low, and that a higher value should be considered.

series literature with the name of system of seemingly unrelated time series equations (SUTSE) model (see for example Harvey, 1989). It is also known that when a SUTSE system contains the same explanatory variables and the system is homogenous[2] each of the N equations can be handled separately. This is a generalization of the well-known result for SURE models which states that when the same variables appear in each equation, the SURE estimator can be obtained by applying Ordinary Least Squares (OLS) to each equation in turn. In the case of the multivariate score test, it is important to notice that once these additional explanatory variables $w_j(\lambda_0)$ have been added to the model, the explanatory variables in each equation are no longer the same and therefore the N equations cannot be handled separately. Therefore, although the score test applied to multivariate time series avoids the introduction of extra parameters in the maximization procedure, it automatically introduces an additional iterative procedure.

3. ROBUST SCORE TEST

The score test, as we have seen in the previous section, avoids the maximization of extra parameters in the likelihood. However, it is not robust to the presence of atypical observations. In order to overcome this problem, in this section we repeatedly fit the forward search algorithm in the way suggested by Atkinson and Riani (2000) and extended to time series by Riani (2004). The algorithm is both efficient and robust. It is efficient because it makes use of the Gaussian likelihood machinery underlying model (18). It is robust because the outliers enter in the last steps of the procedure, and their effect on the statistics of interest is clearly depicted. More generally, this approach allows evaluation of the inferential effect that each time period, either outlying or not, exerts on the fitted model. The key features of the forward search applied to time series can be summarized as follows.

CHOICE OF THE INITIAL SUBSET

We take periods of contiguous observations as the basic sets of our algorithm. These blocks are intended to retain the autocorrelation structure of the whole time series (Cerioli and Riani, 2002). Confining the attention to subsets of contiguous observations ensures that the hyperparameters can be consistently estimated within each block. The initial subset can be obtained through least median or least trimmed of squares applied to these blocks. More precisely, let $v_{t,S_r^{(b+d)}}(\lambda)$ and

---

[2]The system is said to be homogeneous if all linear combinations of its N elements have the same stochastic properties. The simplest example is the multivariate random walk plus noise (Harvey, 1989, p. 435).

$\sigma^2 F_{t,S_r^{(b+d)}}(\lambda)$ be respectively the vector of one step ahead prediction errors and their covariance matrix for time unit t given observations in a subset formed by $b + d$ units $(S_r^{(b+d)})$ where d is associated with the first d observations necessary to initialize the Kalman filter. For a nonstationary time series of order d, twice the profile loglikelihood based on the observations forming $S_r^{(b+d)}$ can be written as:

$$2L_{\max}(\lambda, \theta_{S_r^{(b+d)}}) = \text{const} - \sum_{\{t \in S_r^{(b+d)} \mid t > d\}} \ln |F_t(\lambda)| - Nb \ln \hat{\sigma}^2. \qquad (20)$$

Here $\hat{\sigma}^2$ is estimated using only the observations belonging to the subset excluding the first d, that is, $\hat{\sigma}^2 = \sum_{\{t \in S_r^{(b+d)} \mid t > d\}} v_t'(\lambda) F_t^{-1}(\lambda) v_t(\lambda) / (Nb)$. The symbol $\hat{\theta}_{S_r^{(b+d)}}$ denotes the MLE of the hyperparameters found using only observations belonging to $S_r^{(b+d)}$. Now let $\tilde{v}_{t,S_r^{(b+d)}} = \hat{\sigma}^{-1/2} F_{t,S_r^{(b+d)}}^{-1/2}(\lambda) v_{t,S_r^{(b+d)}}(\lambda)$, $t = d + 1, \ldots, T$ be the vector of one step ahead standardized prediction errors for each unit based on the hyperparameters estimated using observations belonging to $S_r^{(b+d)}$ and let $\tilde{v}_{\{t\},S_r^{(b+d)}}^2$ be the squared tth ordered value. We take as our initial subset of observations the $(b + d)$-tuple which satisfies

$$\min_r \sum_{j=1}^{N} \tilde{v}_{\{[med]\},S_r^{(b+d)}}^2(j) \qquad r = 1, \ldots, k, \qquad (21)$$

where $\tilde{v}_{\{[l]\},S_r^{(b+d)}}^2(j)$ is the jth element of vector $\tilde{v}_{\{[t]\},S_r^{(b+d)}}$, $med = [(T - d)/2]$ and k denotes the number of subsamples which are extracted. Criterion (21) extends the least median of squares method for regression models with independent errors (Rousseeuw, 1984) and univariate response to correlated multivariate observations. In this case, however, standardized residuals instead of raw residuals are considered. In conclusion, we take as our initial subset the stretch of data which minimizes the sum of the medians of the squared one step ahead standardized prediction residuals. It is necessary to remark that 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. (21) is replaced by summation over a fraction of the observations.

Progressing in the Search and Diagnostic Monitoring

The transformed model is repeatedly fitted to subsets of increasing size and are selected in such a way that outliers are included only at the end of the search. For this reason, in each step of size m, we take as the new subset the one formed by the smallest squared Mahalanobis distances

calculated using the standardized one step ahead prediction residuals. For each m the residuals are ordered using observations transformed according to the values specified in λ_0. One major advantage of the forward search over other high-breakdown techniques is that a number of diagnostic measures can be computed and monitored as the algorithm progresses. The focus of this paper is to produce forward plots of the approximate score statistic for testing the significance of the set of constructed variables for different values λ_0, using a separate search for each λ_0. These trajectories of the score tests can be combined in a single picture named the "fan plot" (Atkinson and Riani, 2000). If the number of observations is not large (i.e. less than 200), generally the five most common values of λ_0 $(-1, -0.5, 0, 0.5, 1)$ are sufficient for selecting the appropriate transformation. On the other hand, when the sample size is large we have to consider a finer grid of values of λ_0.

4. SOME EXAMPLES

In order to illustrate the additional insights which come from the application of the suggested procedure, we use two series widely studied in the literature. In the first example (Italian industrial production), the robust score test enables us to quantify the percentage of observations in accordance with the values of each λ_0 and to easily select in a robust way the best value of the transformation parameter. In the second example (kilowatt hours used), the procedure highlights a set of influential values for the choice of the transformation parameter, which was not revealed by the methods based on single deletion diagnostics.

4.1. Italian Industrial Production

Figure 1 shows the series of Italian industrial production, adjusted by the Organisation for Economic Co-operation and Development (OECD) for trading days variations, in the period January 1981–December 2002. This series is characterized by seasonal troughs occurring in August. Usually the ratio between the August value and a month in the spring in the same calendar year is around 0.4 so we can expect that these August figures will strongly influence the choice of transformation. In order to compare our results with previous works we initially consider the period January 1981–December 1996 (Kaiser and Maravall, 1999; Proietti, 2000). Sometimes this series has been analyzed using logs and other times using the original data. The spread level regressions implemented in TRAMO SEATS (Gomez and Maravall, 1996) suggest the log transformation. On the other hand, likelihood-based inferences on the transformation parameter implemented in X12 ARIMA (Findley et al., 1998) suggest that the series should not be transformed. Generally
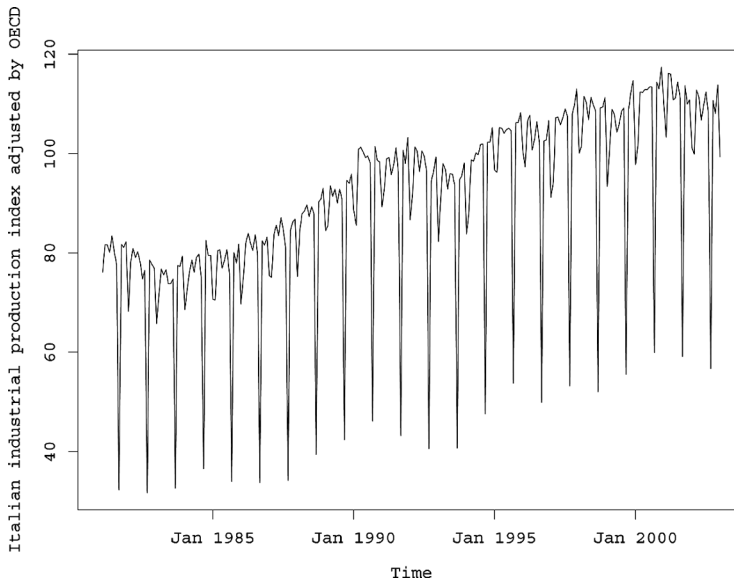
FIGURE 1 Italian industrial production adjusted by OECD over the period January 1981–December 2002. The seasonal troughs occurring in August are evident.

the additive decomposition ($\lambda = 1$) poses no particular feature about the month of August. On the other hand, the multiplicative decomposition seems to flag the months of August as outlying (Proietti, 2000). It is clear, therefore, that the choice of the appropriate transformation turns out to be a relevant issue. Table 1 gives the values of the score test statistic using all the observations for 3 values of λ fitting a basic structural model with monthly trigonometric seasonality (Harvey, 1989, p. 47). As this table shows, both $\lambda = 0.5$ and $\lambda = 1$ give values which are inside the 99% confidence bands of the test. Our purpose is to discover, for example, whether the value of the score test for $\lambda = 1$ is due to the presence of particular observations or is based on all the data.

Figure 2 shows the fan plot (Atkinson and Riani, 2000, p. 89) for three values of λ_0 (0, 0.5, 1). This figure first of all enables us to detect in a robust

TABLE 1 Italian industrial production index over the period January 81–December 96: values of the score test for different values of λ_0 using all the observations

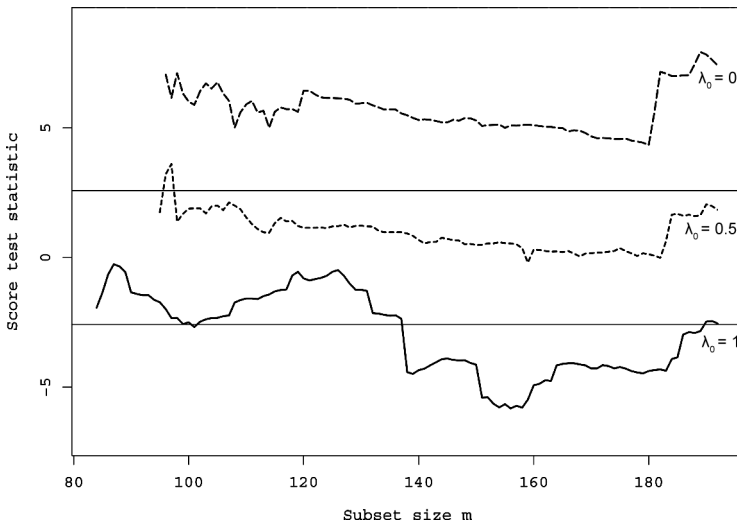| λ_0	$T_{SC}(\lambda_0)$
1	−2.55
0.5	1.84
0	−7.42

FIGURE 2   Fan plot for $\lambda_0 = (0, 0.5, 1)$ for the Italian industrial production index over the period January 1981–December 1986 with 99% theoretical confidence bands coming from the standard normal distribution.

and efficient way the interval of values of λ which can be used to reduce the data to approximate normality. Secondly, it clearly enables us to select the values of the transformation parameters which are supported by the majority of the observations. For example, Figure 2 shows that the score test associated with $\lambda = 1$ lies in the rejection region in the central part of the search. Only the inclusion of the last 3 units which enter the search in this scale brings the value of the test inside the acceptance region. On the other hand, the curve associated with $\lambda = 0.5$ always lies inside the acceptance region throughout the search even if it always takes values slightly greater than zero. There is no doubt therefore that if we limit our analysis to the 5 most common values of λ_0, the square root is the best transformation to obtain approximate normality.

It is interesting now to check how these results change when we consider the whole period January 1981–December 2002 (264 observations). We may expect that with a larger number of observations we have to try values of λ_0 with step 0.25. Figure 3 shows the fan plot for $\lambda_0 = (0.25, 0.5, 0.75, 1)'$. This plot immediately shows that the hypothesis of no transformation is always firmly rejected in the central part of the search. The trajectories associated with $\lambda = 0.5$ and $\lambda = 0.75$ lie inside the confidence region. Note, on the other hand, that if we use single deletion diagnostics, and we delete the most remote observation in the search for $\lambda = 1$ and $\lambda = 0.5$, we would be misled to choose $\lambda = 1$ as the best transformation parameter.
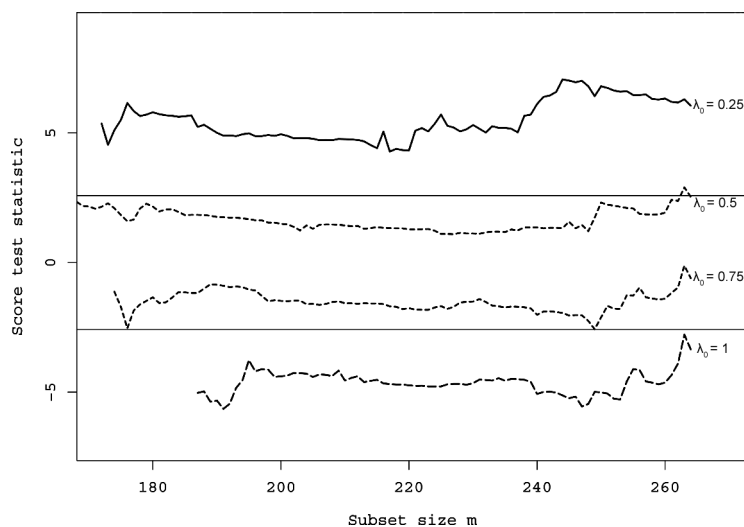
FIGURE 3 Fan plot for $\lambda_0 = (0.25, 0.5, 0.75, 1)$ for the Italian industrial production index over the period Jan 1981–Dec 2002 with 99% theoretical confidence bands coming from the standard normal distribution.

4.2. Kilowatt Hours Used

The second example concerns the series of kilowatt hours used per month in a undefined region over 15 years (180 observations). There are two explanatory variables associated with the weather. The first is the heating degree days, that is the sum over the month of the number of degrees F the average temperature on each day was below 65°F. The second explanatory variable measures the total degree days above 65°F. Certainly, extra power may be expected to be consumed when the weather is cold for heating or when it is hot for cooling. The presence of the regression structure makes it less easy to assess potential difficulties with the fitted model just using raw graphs. This series was introduced by Pankratz (1991) and further analyzed using single deletion diagnostics by Atkinson and Shephard (1996). They both suggest fitting logs. As a matter of fact, the value of the score test when fitting a basic structural model and using all the observations is equal to 0.718. So, it is completely in agreement with the log transformation. The plot of the deletion score test given by Atkinson and Shephard (1996) shows that no particular observation seems to be important for the transformation. Figure 4 shows the fan plot for the four values of $\lambda_0 = (0, 0.25, 0.5, 1)'$. First of all, this plot shows that both $\lambda_0 = 0.5$ (square root transformation) and the $\lambda = 1$ (no transformation) are always rejected. Furthermore, it is also interesting to see that the log transformation is always rejected in the central part of the search (the value of the score is above the upper threshold).
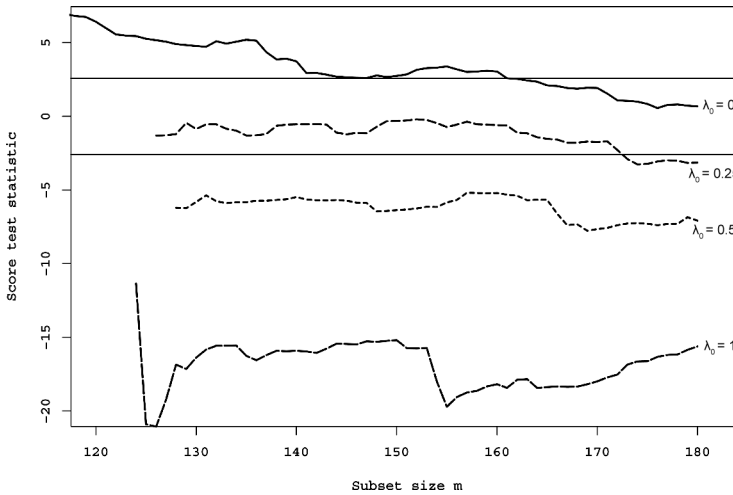
FIGURE 4 Kilowatt data: fan plot. $\lambda = 0.25$ is the best transformation.

The introduction of the last observations brings the value of the test inside the confidence band. On the other hand, the trajectory associated with $\lambda = 0.25$ is always perfectly inside the envelope in the central part of the search and goes below the lower threshold only in the final part. The score test, given that there are no sudden jumps in the trajectories associated with $\lambda_0 = (0,\ 0.25$ and $0.5)'$, shows that no observation is influential for the choice of the transformation if it is considered on its own. The effect appears only when a set of units is considered jointly. Figure 5 shows the monitoring of standardized one-step ahead prediction residuals with the set of influential observations for transformation highlighted. This is a group of observations showing large negative residuals in the central part of the search. At the end, however, their residuals are completely intermingled with those of the other units. It is clear that using single deletion diagnostics it is impossible to detect the effect on the aggregate statistics of this group of observations.

5. SIMULATION ENVELOPES FOR THE FORWARD SCORE TEST

In order to evaluate whether the value of the test was significant throughout the search, we have used the horizontal confidence bands of the standard normal distribution. The question posed in this section is whether this distribution is an appropriate reference distribution. The null distribution of the forward version of the score test for transformation with i.i.d observations has been analyzed by Atkinson and Riani (2002).
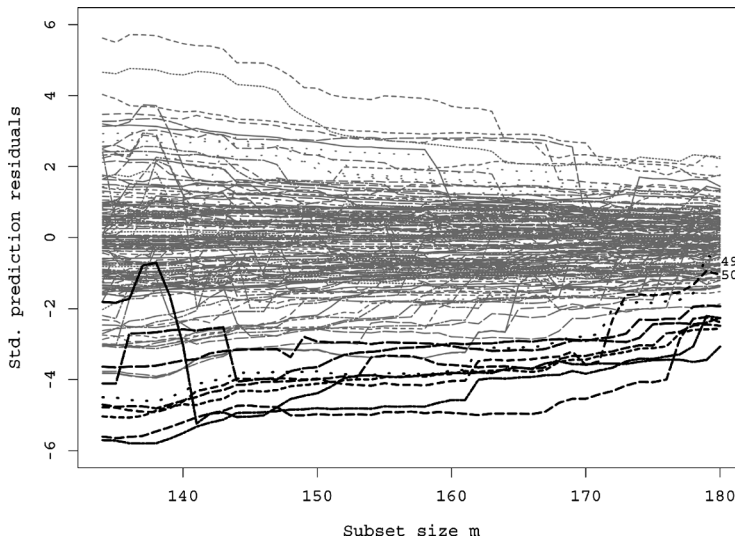
FIGURE 5 Kilowatt data: monitoring of one-step ahead standardized prediction residuals.

These authors conclude that in regression, there is good agreement between the simulated distribution and the normal distribution. However, they noticed that at the end of the search there was evidence of a distribution with slightly longer tails. This phenomenon, which Atkinson and Riani (2002) called the "trumpet effect" was associated with models which showed a low value of the R^2 coefficient. This relationship can be explained by considering the structure of constructed variable plots for simple samples, that is the scatter plots of the residual transformed response against the residual constructed variable. The near parabolic structure of these plots[3] coupled with the fact that the search causes extreme observations on the parabola to enter towards the end of the search explains this phenomenon. Of course, this parabolic structure disappears when there are several explanatory variables and a strong regression structure (high R^2).

The purpose of this section is to analyze whether this remains true with correlated observations. In time series, as is well known, the coefficient of determination does not provide a useful yardstick for assessing goodness of fit. However, two situations provide a benchmark for our study. The first is when the ratio of the variance of the trend and seasonal hyperparameters is very small with respect to the variance of the irregular term (i.e., smaller than 0.001). The second is when the variability of the signal

---

[3]See Atkinson (1985, p. 192) for a Taylor series justification of the parabolic structure and the scatter plots of the residual transformed response against the residual constructed variable.
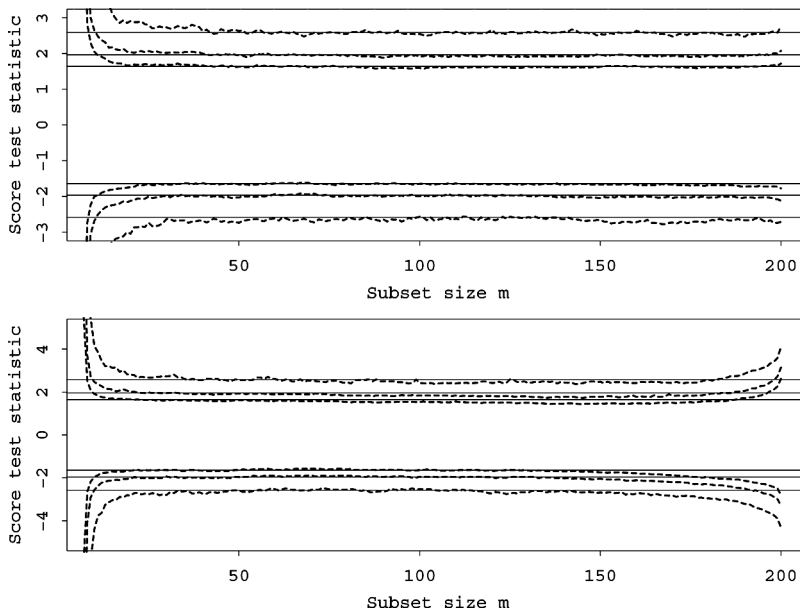
FIGURE 6 Simulation envelopes (dashed bands) and asymptotic percentage points $0.5\%, 2.5\%, 5\%, 95\%, 97.5\%, 99.5\%$ of the standard normal distribution. Upper (lower) panel is associated with high (low) values of signal to noise ratio.

(variance of trend and seasonal) is much larger than the variability of the noise (i.e., greater than 1000). The distribution of the statistic is estimated by simulating 10,000 time series with $T = 200$ observations from a basic structural model and performing one forward search on each. The upper panel of Figure 6 shows the 90%, 95%, and 99% empirical and theoretical confidence bands when the ratio signal to noise is very large. The real distribution of the statistic starts with longer tails than the normal. In the central part of the search, the agreement is extremely good. Finally, in the final part of the search there is evidence of a very slight spreading of the distribution. The lower panel, on the other hand, shows the agreement between these two distributions when the signal to noise ratio is very small. Also, in this case the accordance seems to be very good in the central part of the search. At the end of the search it is possible to notice the so called "trumpet effect." This implies that when the variability of the irregular term is much larger than that of the other components, at the end of the search we have a longer tailed distribution than the normal. In the central part of the search, however, the agreement between real and theoretical envelopes remains good irrespective of the values of the signal to noise ratios.

6. CONCLUSION AND EXTENSIONS FOR FURTHER RESEARCH

There is an appreciable literature on the transformation of time series; see among the others Granger and Newbold (1976), Lenk and Tsai (1990), Tsai and Wu (1992), Pankratz (1991), Guerrero (1993), Atkinson and Shephard (1996). A distinction between this contribution and the previous works on transformations is that the emphasis was on aggregate statistics or on standard deletion diagnostic methods. Using aggregate statistics it is impossible to determine the effect that individual observations exert on a particular transformation. On the other hand, the use of deletion diagnostics is prone to masking and swamping. In our second example, we have shown how the suggested robust procedure is able to detect a group of observations which lead to the choice of a transformation which was not supported by the majority of the data.

ACKNOWLEDGMENTS

REFERENCES

Atkinson, A. C. (1985). *Plots, Transformations, and Regression.* Oxford: Oxford University Press.

Atkinson, A., Shephard, N. (1996). Deletion diagnostics for transformations of time series. *Journal of Forecasting* 5:1–17.

Atkinson, A. C., Riani, M. (2000). *Robust Diagnostic Regression Analysis.* New York: Springer-Verlag.

Atkinson, A. C., Riani, M. (2002). Tests in the fan plot for robust, diagnostic transformations in regression. *Chemometrics and Intelligent Laboratory Systems* 60:87–100.

Atkinson, A. C., Riani, M., Cerioli, A. (2004). *Exploring Multivariate Data with the Forward Search.* New York: Springer-Verlag.

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

Cerioli, A., Riani, M. (2002). Robust methods for the analysis of spatially autocorrelated data. *Statistical Methods and Applications* 11:335–358.

De Jong, P., Chu Chun Lin, S. (1994a). Fast likelihood evaluation and prediction for non stationary state space models. *Biometrika* 81:133–142.

De Jong, P., Chu Chun Lin, S. (1994b). Stationary and nonstationary state space models. *Journal of Time Series Analysis* 15:151–166.

Findley, D., Monsell, B., Bell, W., Otto, M., Chen, B. C. (1998). New capabilities and methods of the x-12-arima seasonal adjustment program. *Journal of Business and Economic Statistics* 16:127–152.

Gomez, V., Maravall, A. (1996). Programs seats and tramo: Instructions for the user. Technical Report 9628, Bank of Spain.

Granger, C. W. J., Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society, Series B* 38:189–203.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting* 12:37–48.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton, New Jersey: Princeton University Press.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Kaiser, R., Maravall, A. (1999). Seasonal outliers in time series. *Estadistica, Journal of the InterAmerican Statistical Institute* 53:97–142.

Koopman, S., Durbin, J. (2003). Filtering and smoothing of state vector for diffuse state-space models. *Journal of Time Series Analysis* 24:85–98.

Lenk, P. J., Tsai, C. L. (1990). Transformations and dynamic linear models. *Journal of Forecasting* 9:219–232.

Pankratz, A. (1991). *Forecasting with Dynamic Regression Models*. New York: Wiley.

Proietti, T. (2000). The seasonal adjustment of the Italian industrial production series. In: *Seasonal Adjustment Procedures: Experiences and Prespectives*. Annali di Statistica 129. Rome: Istituto Nazionale di Statistica, pp. 373–378.

Riani, M. (2004). Extension of the forward search to time series. *Studies in Non Linear Dynamics and Econometrics* 8:1–23.

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

Tsai, C.-L., Wu, X. (1992). Transformation model diagnostics. *Technometrics* 34:197–202.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias. *Journal of the American Statistical Association* 57:348–368.