Studies in Nonlinear Dynamics & Econometrics

Volume 8, *Issue* 2 2004 *Article* 2

LINEAR AND NONLINEAR DYNAMICS IN TIME SERIES
ESTELLA BEE DAGUM AND TOMMASO PROIETTI, EDITORS

Extensions of the Forward Search to Time Series

Marco Riani*

*University of Parma, mriani@unipr.it

Extensions of the Forward Search to Time Series*

Marco Riani

Abstract

This paper extends the forward search technique to the analysis of structural time series data. It provides a series of powerful new forward plots that use information from the whole sample to display the effect of each observation on a wide variety of aspects of the fitted model and shows how the forward search, free from masking and swamping problems, can detect the main underlying features of the series under study (masked multiple outliers, level shifts or transitory changes). The effectiveness of the suggested approach is shown through the analysis of real and simulated data.

*Marco Riani Department of Economics, University of Parma Via Kennedy, 6 43100 Parma Italy
e-mail:mriani@unipr.it

1 Introduction

The forward search (FS) is a powerful general method for detecting multiple masked outliers, for determining their effects on models fitted to data and for detecting systematic model inadequacy. This method was originally introduced for models which assumed independent observations: linear and non linear regression (Atkinson and Riani, 2000, 2002), generalized linear models (Atkinson and Riani, 2001) and multivariate analysis (Riani and Atkinson, 2001). In this paper we extend the forward search technique to the analysis of structural time series data. The basic ingredients of the FS are a robust start from an outlier-free subset of observations, a criterion for progressing in the search, which allows the subset to increase in size by one or more observations at each step, and a set of diagnostic tools that are monitored along the search. The robustness of the FS stems from the very definition of its algorithm, starting from "good" data points and including outliers at the end of the procedure. Computation of high-breakdown estimators (e.g. Hampel, *et al.*, 1986; Rousseeuw and Basset, 1991) is not required, except possibly at the starting stage. Indeed, the application of efficient likelihood or moment-based methods at subsequent steps of the FS provides the analyst with more powerful tools than those obtained via traditional high-breakdown estimation. The flexibility of the FS makes this procedure suited for extensions to areas other than multiple regression and multivariate analysis. This is especially true in time series, where it is often necessary to detect and model sudden or unexpected events. It is well known that outliers or structural changes in the observed time series may seriously damage identification and estimation of the suggested ARIMA or structural model (e.g. Koopman and Harvey, 1992; Chen and Liu, 1993), because such occurrences can introduce serious bias in the sample autocorrelation function. In time series the standard procedures for automatic outlier detection and correction consider four types of outliers, namely: additive (AO), innovational (IO), level shift (LS), and transitory change (TC) (Tsay, 1986; Chen and Liu, 1993; De Jong and Penzer, 1998). The AO represents a single spurious observation, the IO a pulse shock to the noise sequence which propagates to the observed time series, the LS a step function and the TC a spike that takes a few periods to disappear.

The main drawback of the traditional procedures for detecting outliers in time series is that they start with the specification of a model for the observed series as if there were no outliers. Given that each kind of outlier is supposed to be generated by an ARIMA formulation, the model fitted to all the observations is refitted under the null assumption there is an outlier at time t. This is done for each type of outlier assuming that each observation in turn y_t, $t = 1, \ldots, T$ is an atypical observation. In other words, the ARIMA filter $\Pi = \phi(B)/\theta(B)$ based on the parameters estimated using all the observations, is repeatedly used $T \times 4$ times to obtain a statistic λ_t for each observation and each type of outlier: AO, IO, LS and TC (e.g Tsay, 1986 or Gomez and Maravall, 1994). The maximum value of λ_t is compared with a predetermined critical value C to decide whether a potential outlier is present in the time series. Once an outlier is found, the filter Π is revised and reapplied for each observation and for each type of outlier in an iterative way. These procedures do not detect the k outliers at once, but proceed in several iterations detecting them one by one. In order to avoid these problems and to be able to detect stretches of over influential observations, Bruce and Martin (1989) suggested leave-k-out diagnostics. The parameters of the model fitted to the full data set are compared with those generated by fitting the model to the data when a stretch of k points are taken to be missing. This approach has been extended to state space models by Proietti (2003). This author shows in an elegant way how to use a reverse run of the Kalman filter on the smoothing error to compute leave-k-out diagnostics. These methods however, become computationally infeasible when k is large and the points are not consecutive, and may still suffer from masking, swamping or smearing effects if the number of outliers is greater than k. Finally, these approaches start with estimated parameters based on all T observations, as if no outlier were present in the data. So, the methods will be able to discover the real structure of the data only if the joint effect of the atypical observations does not destroy the validity of the estimate of the covariance structure of the data.

Beyond the problem of the detection of atypical observations, another important issue concerns the effect of individual observations on the maximum likelihood estimate of the parameters. In other words, because of the way in which models are fitted, we lose information about the effect of individual observations on inferences about the form and parameters of the model. For example, it is very important to know whether a component (trend, seasonal or cycle) becomes deterministic, if we exclude certain observations from the estimation process. Similarly, it is useful to investigate the stability of the estimate of the period of the suggested stochastic cycle. More generally: it is useful to know which estimates of the hyperparameters are stable and which are those affected by particular observations. In this paper we provide a series of powerful new forward plots that use information from the whole sample to display the effect of each observation on a wide variety of aspects of the fitted model and we show how the forward search can detect the main underlying features of the series under study (masked multiple outliers, level shifts or transitory changes).

The structure of the paper is as follows. In section 2, given that in our approach we repeatedly use the diffuse Kalman filter to estimate the hyperparameters in each step of the search, we briefly review the state-space formulation and give some details about the way we initialize the Kalman filter. In section 3 we show how the forward search routines can be extended to the analysis of time series data using the diffuse Kalman filter with missing observations. In section 4 we apply the suggested procedure to real and simulated time series. Section 5 contains concluding remarks and directions for future research.

2 Non stationary state space form

As is well known, the state-space form and the Kalman filter provide a unifying tool for state-space model likelihood evaluation and prediction. Both ARIMA and structural time series models can be put into the state space formulation and can be estimated using the Kalman filter. The vector time series y_t ($t = 1, 2, \ldots, T$), with N elements is said to be generated by a state space model if

$$y_t = Z_t\alpha_t + X_t\beta_t + G_t\epsilon_t \quad (2.1)$$

$$\alpha_{t+1} = T_t\alpha_t + W_t\beta_t + H_t\epsilon_t \quad (2.2)$$

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. If vector β_t in equations (2.1) and (2.2) 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 $y_1, \ldots, y_{t-1}$

$$a_{t|t-1} = E(\alpha_t | y_1, \ldots, y_{t-1})$$

and its mean squared error (MSE)

$$MSE(a_{t|t-1}) = E[(a_{t|t-1} - \alpha_t)(a_{t|t-1} - \alpha_t)' | y_1, \ldots, y_{t-1}] = P_{t|t-1}.$$

The Kalman filter is the set of recursions

$$v_t = y_t - Z_t a_{t|t-1} \qquad F_t = Z_t P_{t|t-1} Z_t' + G_t G_t' \tag{2.3}$$

$$q_t = q_{t-1} + v_t' F_t^{-1} v_t \qquad K_t = (T_t P_{t|t-1} Z_t' + H_t G_t') F_t^{-1} \tag{2.4}$$

$$a_{t+1|t} = T_t a_{t|t-1} + K_t v_t \qquad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t' - K_t F_t K_t' \tag{2.5}$$

with $q_0 = 0$. The filter innovations (one step ahead prediction errors) are indicated with v_t and their variance with $F_t = var(v_t) = var\{y_t - E(y_t | y_1, \ldots, y_{t-1})\}$. These two quantities form the necessary ingredients for the computation of the loglikelihood

$$l(\theta) = -0.5NT \ln \sigma^2 + \sum_{t=1}^{T} \ln |F_t| + \sigma^{-2} q_T. \tag{2.6}$$

The maximum likelihood estimate of σ^2 in equation (2.6) is

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

Vector β_t in equations (2.1) and (2.2) may contain random, fixed or diffuse effects. An element is said to be diffuse if its variance tends to infinity. The diffuse assumption reflects parameter uncertainty and needs to be applied every time a time series model is nonstationary. In this last case the classical approach (e.g. Harvey, 1989) suggests starting the filter with $P_{1|0} = \kappa I$, with κ equal to a large but finite number. This approximation, however, especially in the context of the forward search which requires a high number of computations, produces noticeable numerical problems. In order to avoid numerical errors in the initialization of the Kalman filter, we used the approach suggested by De Jong and Chu Chun Lin (1994a, 1994b) as follows. If d is the number of nonstationary elements and k is the number of explanatory variables, up to observation y_{d+k} we run the so called diffuse Kalman filter which is given by the following recursions

$$V_t = (0, y_t) - Z_t A_{t|t-1} \qquad F_t = Z_t P_{t|t-1} Z_t' + G_t G_t' \tag{2.7}$$

$$Q_t = Q_{t-1} + V_t' F_t^{-1} V_t \qquad K_t = (T_t P_{t|t-1} Z_t' + H_t G_t') F_t^{-1} \tag{2.8}$$

$$A_{t+1|t} = T_t A_{t|t-1} + K_t V_t \qquad P_{t+1|t} = T_t P_{t|t-1} T_t' + H_t H_t' - K_t F_t K_t'. \tag{2.9}$$

Matrix $A_{t|t-1}$ can be partitioned as $(A^{\dagger}_{t|t-1} a^0_{t|t-1})$ and has $d + k + 1$ columns; $a^0_{t|t-1}$ (the last column of $A_{t|t-1}$) is the standard Kalman filter recursion initialized with the variance covariance matrix of the non-stationary elements set equal to zero. The first $k+d$ columns of $A_{t|t-1}$ ($A^{\dagger}_{t|t-1}$) are obtained by applying the same Kalman filter to the zero observations which make up the first $d+k$ columns of matrix $(0, y_t)$. In the calculation used in this paper we used the diffuse Kalman filter up to step $d+k$ and then collapse it to the usual Kalman filter. The collapse is realized using the appropriate blocks of the $(d+k+1)\times(d+k+1)$ matrix Q_{d+k}. More precisely

$$Q_{d+k} = \begin{pmatrix} q^0_{d+k} & s'_{d+k} \\ s_{d+k} & Q^\dagger_{d+k} \end{pmatrix}$$

where $Q^\dagger_{d+k}$ has dimension $(d+k)\times(d+k)$. The appropriate starting values can be obtained from the output of the diffuse Kalman filter as[1]

$$a_{d+k+1|d+k} = a^0_{d+k+1|d+k} - A^\dagger_{d+k+1|d+k}\left(Q^\dagger_{d+k}\right)^{-1} s_{d+k} \quad (2.10)$$

$$P_{d+k+1|d+k} = P^0_{d+k+1|d+k} + A^\dagger_{d+k+1|d+k}\left(Q^\dagger_{d+k}\right)^{-1} A^{\dagger'}_{d+k+1|d+k}. \quad (2.11)$$

The resulting likelihood becomes

$$l(\theta) = -0.5N(T-d-k)\ln\sigma^2 + \sum_{t=d+k+1}^{T} \ln|F_t| + \sigma^{-2} q_T$$

with $q_{d+k} = 0$.

When a set of observations is missing, the vector v_t and the matrix K_t are set to zero for these values $(t = d+1, \ldots, T)$, that is $v_t = 0$ and $K_t = 0$ (e.g. Harvey and Pierse, 1984) and the Kalman updates become

$$a_{t+1|t} = T_t a_{t|t-1} \qquad P_{t+1|t} = T_t P_{t|t-1} T'_t + H_t H'_t. \quad (2.12)$$

This simple treatment of missing observations makes the state space approach ideal for the forward search algorithm which requires repeated estimation on subsets of data non necessarily contiguous.

3 The forward search in time series

The forward search is made up of three steps: choice of the initial subset, progressing in the search and diagnostic monitoring. In the following subsections we will examine in detail these three steps.

3.1 Choice of the initial subset

Details of the forward search for regression on a single response variable are given in Atkinson and Riani (2000). The method starts by fitting a small, robustly chosen, subset of m of the n observations to the data. In time series the initial subset can be chosen among k blocks of contiguous observations of fixed dimension b. The idea of block sampling is to retain the same dependence structure as the original data set (e.g. Haegerty and Lumley, 2000). To find

[1] In equations (2.10) and (2.11) we supposed that $Q^\dagger_{d+k}$ is invertible. If this is not the case we can collapse the diffuse Kalman filter to the usual Kalman filter in the first step in which this matrix becomes invertible.

the initial subset we perform an exhaustive search of all possible blocks and choose the one which satisfies a least median of squares criterion. Specifically, if the suggested time series model is non-stationary of order d (without loss of generality we suppose there are no regression effects), we can divide the last $(T-d)$ observations of the time series into k subsamples, each made up of the first d initial data $y_1, \dots, y_d$ plus a set of $[(T-d)/k]$ contiguous observations where $[\cdot]$ is the integer part of a number. If $(T-d)/k$ is an integer we have exactly k subsamples. In the general case we have $k+1$ subsamples, the first k of size $d+[(T-d)/k]$, and the last of size $T-[(T-d)/k]k$. Without loss of generality in this paper we assume that $(T-d)/k = b$ (say) is an integer. The units forming subsample $r = 1, 2, \dots, k$, say $S_r^{(b+d)}$, are $y_1, \dots, y_d, y_{d+1+(r-1)b}, \dots, y_{d+br}$.

Another possibility to perform block sampling would be to extract without replacement l observations from the last $T-d$. If for example observation u is selected we extract all observations from u to $u+r$. In order to form the initial subset we delete the eventual duplicate units.

Now, let $v_{t,S_r^{(b+d)}}$ and $F_{t,S_r^{(b+d)}}$ be respectively the vector of one step ahead prediction errors and their covariance matrix for time unit t given observations in $S_r^{(b+d)}$. For a non-stationary time series of order d, generalizing equation (2.6), minus twice the log of the normal likelihood based on the observations forming $S_r^{(b+d)}$ can be written as

$$l(\theta_{S_r^{(b+d)}}) = -0.5Nb \ln \sigma^2 + \sum_{t=d+1+(r-1)b}^{d+rb} v'_{t,S_r^{(b+d)}} F^{-1}_{t,S_r^{(b+d)}} v_{t,S_r^{(b+d)}}. \qquad (3.1)$$

Here σ^2 is estimated using $\hat{\sigma}^2_{S_r^{(b+d)}} = \sum_{t=d+1+(r-1)b}^{d+rb} v'_{t,S_r^{(b+d)}} F^{-1}_{t,S_r^{(b+d)}} v_{t,S_r^{(b+d)}}/(Nb)$. 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 $\tilde{v}_{t,S_r^{(b+d)}} = \hat{\sigma}^{-1/2}_{S_r^{(b+d)}} F^{-1/2}_{t,S_r^{(b+d)}} v_{t,S_r^{(b+d)}}$, $t = d+1, \dots, T$ 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 $\tilde{v}_{\{t\},S_r^{(b+d)}}$ be the t-th ordered residual. We take as our initial subset of observations the $(b+d)$-tuple which satisfies

$$\min_r \sum_{j=1}^{N} \tilde{v}^2_{\{[med]\},S_r^{(b+d)}}(j) \qquad r = 1, \dots, k \qquad (3.2)$$

where $\tilde{v}^2_{\{[t]\},S_r^{(b+d)}}(j)$ is the jth element of vector $\tilde{v}_{\{[t]\},S_r^{(b+d)}}$ and $med = [(T-d)/2]$.

Criterion (3.2) 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. As the number of k subsamples considered increases, the probability of having at least an initial subset of data which does not contain outliers increases. On the other hand, if the size of the initial subsets is too small, the estimates of the hyperparameters can be unstable. We suggest a value of $b \approx \sqrt{T}$ to ensure a balance between the statistical properties of the estimated hyperparameters and the robustness of the method. As we see in the examples the choice of the initial subset is not crucial, because in the majority of cases the final part of the search (which is certainly the most interesting) is unaffected by this choice.

3.2 Progressing in the search

Given a subset of dimension $m \geq b + d$, $m = b + d, \ldots, T - 1$, the FS in time series moves to dimension $m + 1$ by selecting the $m + 1 - d$ observations with the smallest squared standardized prediction residuals, the observations being chosen by ordering all $(T - d)$ squared residuals $\tilde{v}^2_{t,S^{(m)}} = v'_{t,S^{(m)}} F^{-1}_{t,S^{(m)}} v_{t,S^{(m)}}$, $t = d + 1, \ldots, T$. In order to initialize the Kalman filter the first d observations are always kept in the subset at each step of the forward search.

More precisely at step m the likelihood which is maximized is given by

$$l(\theta_{S^{(m)}}) = -0.5N(m - d) \ln \sigma^2 + \sum_{t \in S^{(m)}} v'_{t,S^{(m)}} F^{-1}_{t,S^{(m)}} v_{t,S^{(m)}}. \tag{3.3}$$

The summation in equation (3.3) includes all the observations in the subset excluding the first d of the time series. Similarly, σ^2 is estimated using only the units forming $S^{(m)}$. It is important to distinguish the residuals used for plotting and those used inside the Kalman filter. The units not included in the subset are treated as missing, so for these observations $K = 0$ and $v = 0$ and the update is skipped. However, for plotting and progressing in the search the one step ahead prediction values v_t and their covariance matrix F_t are computed for all units. So, given a subset of size m, we have $(T-d)N$ standardized prediction residuals.

Similarly to what happens in multivariate analysis when the Mahalanobis distances are monitored, if the subset size is small the units not included will have a very high Mahalanobis distance which tends to decrease during the forward search. In order to have stable curves throughout the search Atkinson and Riani (2000) in each step scale the residuals with the final estimate of σ based on n observations. In time series if we want stable trajectories we may scale standardized residuals in each step with the square root of the generalized prediction error variance (see for example Harvey, 1989; p. 445) based on all observations. Such a rescaling increases emphasis on the later parts of forward plots and produces more stable trajectories (Atkinson, Riani and Cerioli, 2003).

As concerns the computational cost of the forward search, if we take $b = \sqrt{T}$, in order to find the initial subset we must fit $\sqrt{T}$ models each based on $\sqrt{T}$ observations. Given we have $T - \sqrt{T}$ updates, the parameters are reestimated $\sqrt{T} + T - \sqrt{T} = T$ times. In traditional backwards methods, if each time we reestimate the parameters and we stop at step $T - k$, for each type of outlier the model must be recomputed $(k + 1) \times T$ times. Finally, if we consider multiple deletion and reestimation of the parameters, the model must be recomputed $\sum_{j=0}^{k} \binom{T}{j}$ times. In addition, the backwards approach may fail to reveal the real structure of the data due to the well known masking and swamping problems.

3.3 Diagnostic monitoring

A major advantage of the FS approach is to provide the user with a number of informative pictures displaying all the diagnostics computed along the search. For instance, the anomalous behaviour of outliers or level shifts is clearly revealed by the individual trajectories in the forward plot of standardized one step ahead prediction errors, even when standard deletion diagnostics suffer from masking. An additional bonus of the forward search algorithm is that it is often possible to rank the observations according to their entrance step into $S^{(m)}$. This gives an ordering of the data according to their degree of agreement with the null model, with observations furthest from it joining $S^{(m)}$ in the last steps of the procedure. Finally, analyzing the individual trajectories of the units along the FS, we can detect the relationships among the observations: this helps to distinguish between the different types of outliers and interventions.

As in regression (Atkinson and Riani, 2000; p. 34), we found it useful to monitor the maximum standardized prediction residual for the units included in the subset and the minimum for those which are outside. While the plot of the maximum is characterized by an upward jump when the first outlier is included in the subset, the plot of minimum shows an upward jump in the step prior to the inclusion of the first atypical observation in the subset.

4 Some numerical work

As an illustration of the suggested approach, in this section we apply the forward search to some real and contaminated time series. In section 4.1 we consider the monthly Italian Industrial Production Index (IPI) from January 1981 to December 1996. We have used this series because these data have already been analyzed with different purposes by various authors (e.g. Kaiser and Maravall, 2001). In section 4.2 we see how the suggested approach responds to an artificial temporary level shift. In section 4.3 we analyze the series of US index of industrial production for the textile sector.

4.1 Italian industrial production index

The purpose of this section is to show in practice the additional gain of information which comes from the forward search. In this section we consider two versions of IPI, the series given by OECD which is adjusted for calendar effects (from now on OECD series) and the original series. Figure 1 shows that this series exhibits a slight increasing trend and a strong seasonal pattern with sharp peaks during the month of August. As expected, the variability of the seasonal peaks seems much more pronounced in the original series (right panel of Figure 1).

We start the analysis applying the basic structural model (Harvey, 1989) with trigonometric seasonality to the OECD series. The basic structural model can be expressed as

$$y_t = \mu_t + \gamma_t + \epsilon_t$$

where μ_t is the level of the trend component, γ_t is the seasonal component and the irregular ϵ_t, is NID$(0, \sigma^2_\epsilon)$. The general formulation for the trend is locally

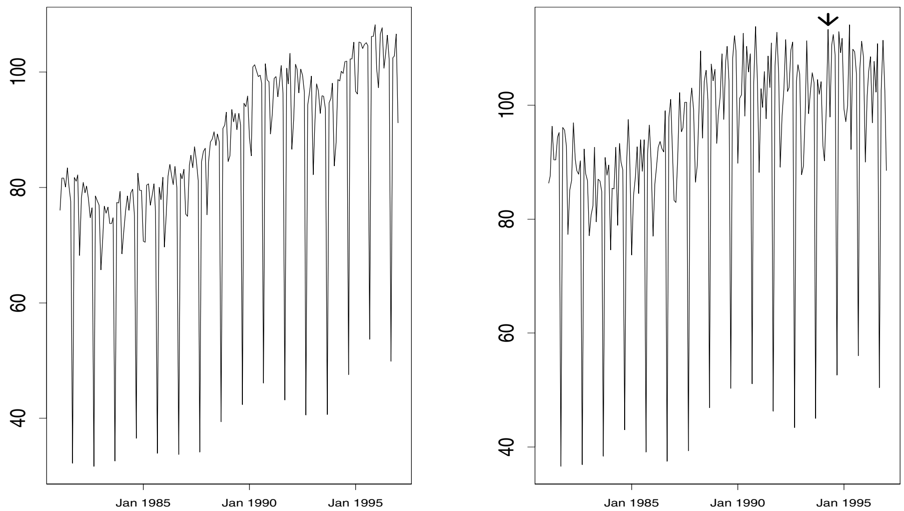

Figure 1: Left panel Italian industrial production adjusted by OECD. Right panel: original Italian industrial production series

linear

$$\begin{aligned}\mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t,\end{aligned}$$

where η_t and ζ_t are mutually uncorrelated white noise disturbances with zero mean and variances $\sigma^2_\eta = q_\eta \sigma^2_\epsilon$ and $\sigma^2_\zeta = q_\zeta \sigma^2_\epsilon$. As it is well known, the effect of η_t is to allow the level of the trend to shift up and down, while ζ_t allows the slope to change. The larger the variances q_η and q_ζ, the greater the stochastic movements in the trend relative to the variability of the irregular component ϵ_t. The seasonal pattern γ_t is usually modelled as the sum of $[s/2]$ cyclical components allowed to vary over time with common variance $\sigma^2_\omega = q_\omega \sigma^2_\epsilon$. This model can easily be put in the state form described in equations (2.1) and (2.2), (see for example Durbin and Koopman, 2001). Given that the basic structural model for monthly time series has 13 non-stationary elements, we run the diffuse Kalman filter up to step $m = 13$ to find appropriate starting values for the remaining $T - 13$ observations.

Figure 2 is a classical representation used in forward search analysis to monitor residuals. It shows the one step ahead standardized prediction errors for each unit in the last 60 steps. For each m the residual have been scaled with the estimate of σ based on $S^{(m)}$ but using all $(T - d)$ observations. Each line refers to a time period. The plot shows that throughout the search (with the exception of observation 140 - August 1992) all cases have standardized prediction residuals inside the 99% asymptotic confidence bands (-2.58, 2.58). This plot clearly shows that there are no masked outliers in the series.

Another question of interest the FS easily enables to answer concerns the stability of the estimates of the hyperparameters. In this example the final values when $m = n$ are $\hat{q}_\eta = \hat{\sigma}^2_\eta/\hat{\sigma}^2_\epsilon = 0.1783$, $\hat{q}_\zeta = \hat{\sigma}^2_\zeta/\hat{\sigma}^2_\epsilon = 0.0030$ and $\hat{q}_\omega = \hat{\sigma}^2_\omega/\hat{\sigma}^2_\epsilon = 0.0046$. The typical questions are: what are the most influential

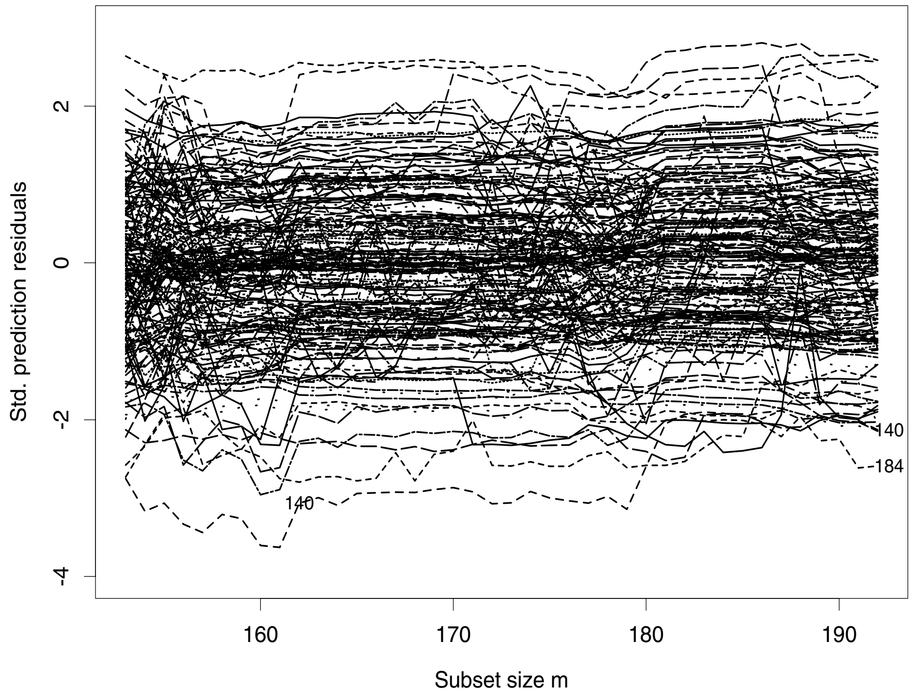

Figure 2: Italian industrial production index corrected by OECD: monitoring of the one step ahead standardized prediction residuals

observations in the estimate of q_η, q_ζ or q_ω? Is the stochastic seasonality diffused throughout the data or is it due to the presence of particular observations? Figure 3, which shows the monitoring of estimates of the q ratios along the last 50 steps of the forward search, enables us to answer all these questions. The left panel shows that the estimate of q_η seems to fluctuate in the interval 0.1 and 0.7 but is always greater than 0 throughout the search. The center panel shows that the variability of the slope movements seem negligible with respect to the irregular in steps 181-186. The inclusion of the last 6 units seems to bring the estimate to values similar to those seen in steps 160-180. A benefit from the forward search is that it enables to link the effect of each unit on the estimate of the hyperparameters. If the model is correctly specified the trajectories of the estimates of the hyperparamters should fluctuate around a certain threshold without particular pattern. This is precisely what happens in this example.

We conclude the analysis of this series with two plots. Figure 4 shows the monitoring of squared maximum standardized residual among the units belonging to the subset (left panel) and the squared minimum standardized one step ahead prediction errors among the observations not included in the subset (right panel). These two plots are very useful for outlier detection, because the first atypical observation which is included in the subset generally causes an upward jump in the plot which monitors the maximum residual in the subset. On the other hand, the plot which monitors the minimum residuals shows an upward jump in the step prior to the inclusion of the first outlier. Figure 4 shows that these two curves increase smoothly and that there are not sudden jumps. The conclusion is that in the OECD series there are no masked outliers. In this case,

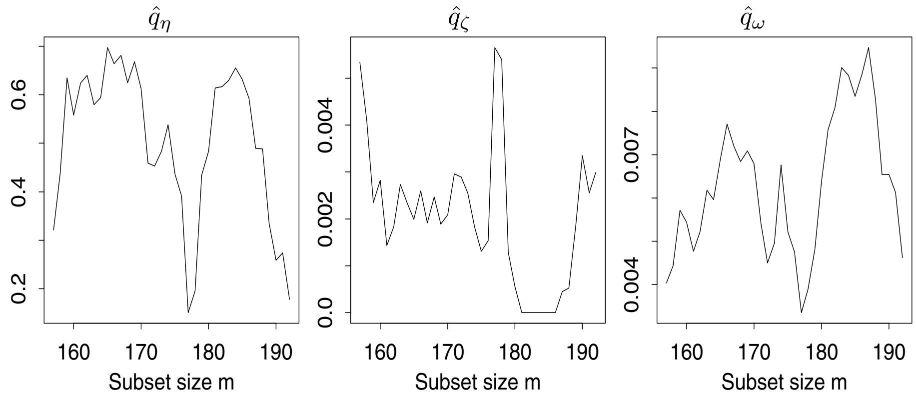

Figure 3: Italian industrial production index corrected by OECD: monitoring of the estimates of the hyperparameters

the FS simply provides an ordering of the data according to their agreement with the suggested model.

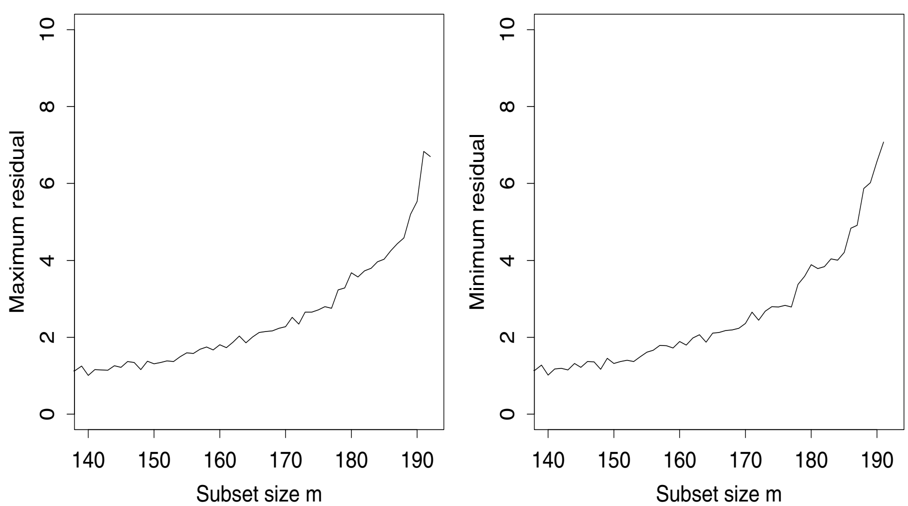

Figure 4: Italian industrial production index corrected by OECD: monitoring of the maximum squared standardized residual among the units belonging to the subset and minimum squared standardized residual among the units outside the subset

Let us now move to the analysis of the original industrial production index. Figure 5 shows the monitoring of squared maximum and minimum standardized residuals for this new series and must be compared with Figure 4. We can see that the last three units which enter the forward search (140, 145, 172) cause a sudden change in the slope of the two curves. In fact the values for these 3 units have been strongly modified by OECD. For example, the value of the production index for April 1995 (unit 172) passes from 104.1 to 92.2. Another plot which is useful to detect areas of misspecification in the model is the so called "entry plot" (Figure 6). Dots indicate the presence of an observation in

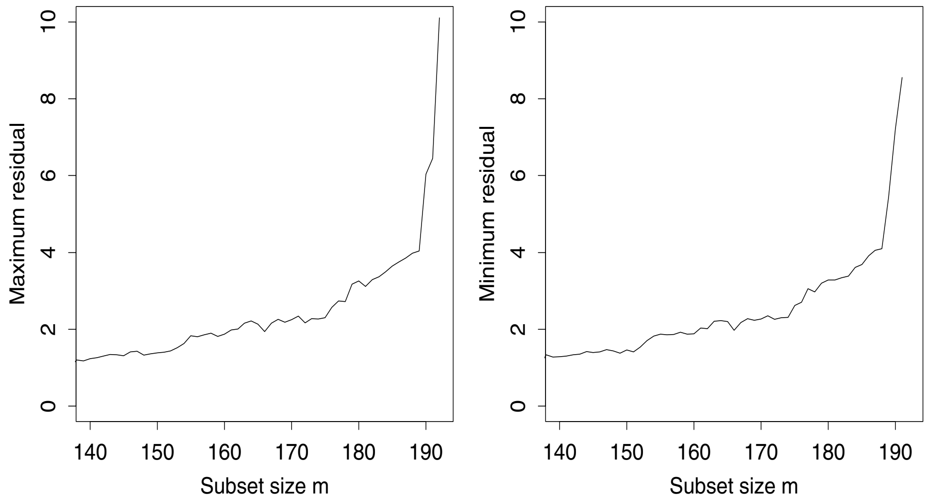

Figure 5: Italian industrial production index: monitoring of the maximum squared standardized residual among the units belonging to the subset and minimum squared standardized residual among the units outside the subset

the subset, so that the number of dots increases towards the right of the graph as the subset size does. This plot is particularly useful, as we will see, when there is a structural break or a level shift, because the set of units located just after the structural break will tend to enter the subset together and the entry plot will be characterized by the presence of white space for certain periods of time. Figure 6 shows a clear strip of which space associated with the period March 1994 - December 1994 (observations 159-168). While units 163 and 165 join the subset when m is around 155, all the others of this group enter the subset in the final part of the search. This suggests that this subgroup of observations is not in agreement with the rest of the model. The trajectories of the residuals for these units together with unit 172 (the last to enter the forward search) are shown in Figure 7. This plot shows that the trajectories of the residuals of this group of units is very similar throughout the forward search and that in the central part they always show a standardized residual around 2. Note that in the final part of the search, due to the well known masking effect, they show small residuals. Figure 8 shows the standardized one step ahead prediction residuals in the last step of the forward search. The units which have a value greater than 2 in absolute value are 41, 64, 140, 145 and 172. The group of units associated with the period March 1994 - December 1994 (observations 159-168) does not show anything particular.

This example shows that going backwards it would be difficult to detect the lack of fit for this group of units.

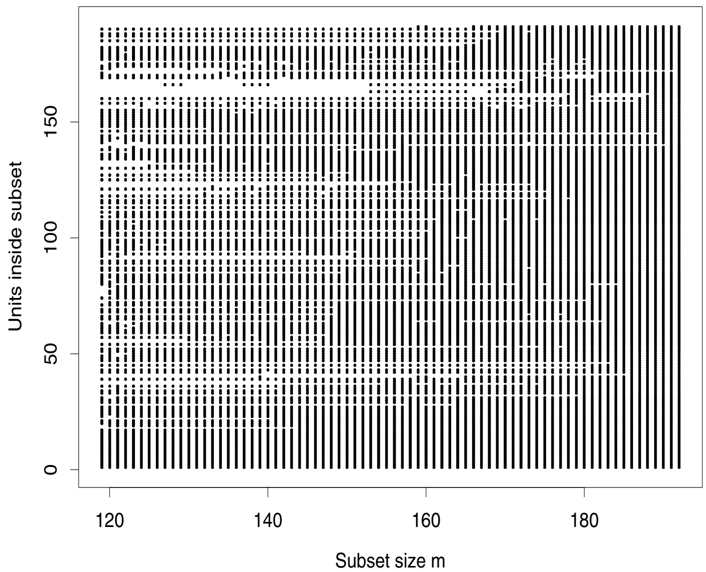

Figure 6: Italian industrial production index: monitoring of units belonging to the subset in final part of the forward search (entry plot)

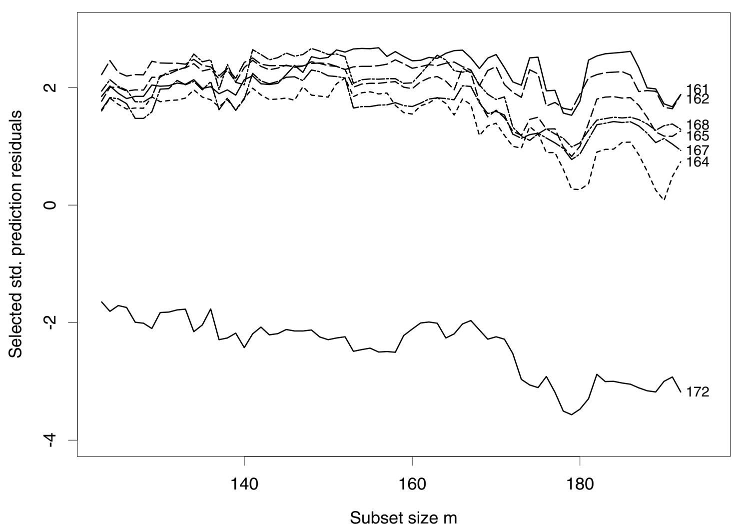

Figure 7: Italian industrial production index: monitoring of selected one step ahead standardized prediction residuals

If we look carefully at the right hand panel of Figure 1, it is clear that the months from March 1994 have values of the production index generally higher than the previous ones suggesting a slight break in the series. The beginning of this break has been marked with a downwards arrow in the right panel of Figure 1. Once our attention has been drawn to that period of time this feature appears clear in the plot of the data. Note that this "break" does not appear at all in the OECD adjusted series (left hand panel of Figure 1). The entry plot for the OECD adjusted series, here not given, is devoid of white strips. Finally, Figure 7 shows that the trajectory of unit 172 seems to be specular to that of units 160-169. In other words, the residual for unit 172 seems to increase in absolute value when the residuals associated with the group of units 160-169 seem to decrease. While traditional procedures for outlier detection simply point out that unit 172 has a high negative residual, the forward search enables to show the connection between different units and different periods of time. This is an example with real data of the amount of additional extra information about the fitted model and the relationships among the units the forward search provides.

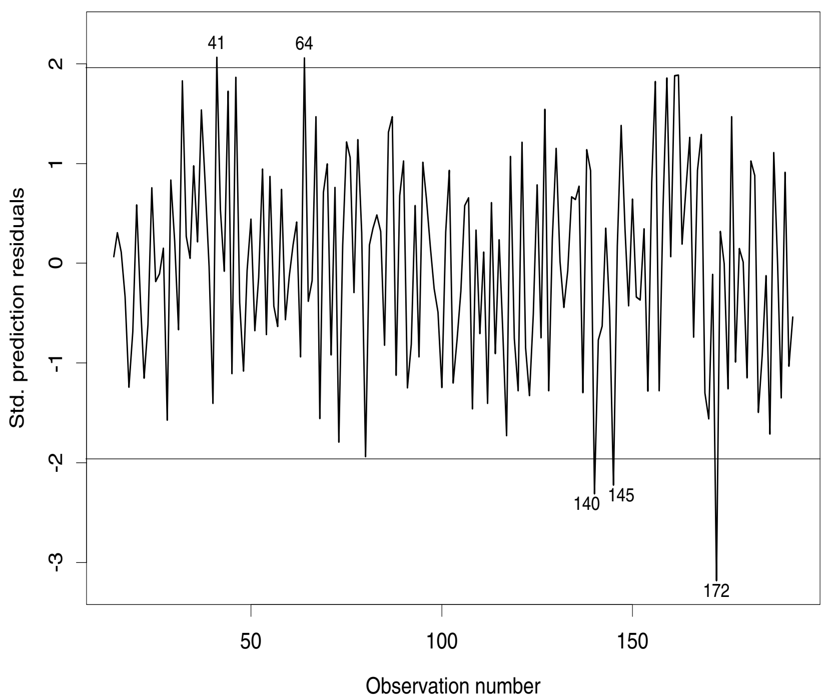

Figure 8: Italian industrial production index: standardized one step ahead prediction residuals in the last step of the forward search

4.2 Italian industrial production index: temporary level shift

In this section we explore how our method reacts to contaminated data. We create a temporary level shift in the data analyzed in the last section from observation 160 (April 1994) to 169 (January 1995). The original values are inflated by 20%. Figure 9, shows the monitoring of one step ahead standardized prediction residuals from step $m = 75$. The contaminated units are clearly visible at the top of the plot as the search progresses.

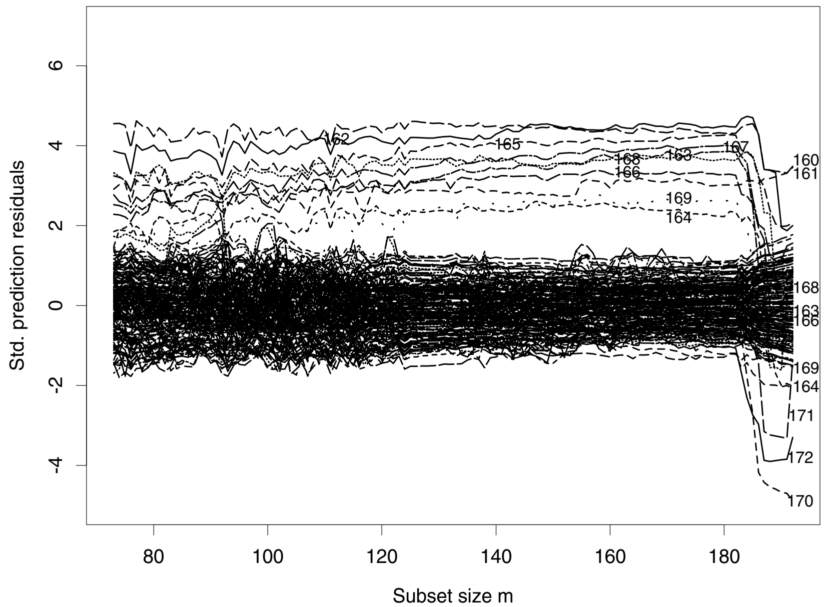

Figure 9: Italian industrial production index with temporary level shift: monitoring of the one step ahead standardized prediction residuals

The use of residuals from a robust starting point in this case is not sufficient to detect the stretch of atypical observations. Finally, it is interesting to notice the huge masking effect at the end of the search. When $m = n = 192$ the observations with the highest residuals in absolute value are 170 and 172. The monitoring of residuals shows that these two observations always have a residual in agreement with the rest of the observations throughout all the search. Table 1, which reports the units included and removed from the subset in the last 12 steps of the forward search, points out that when $m = 186$, $m = 187$ and $m = 188$, two units (instead of one) enter the subset at the same time and units 170 and 172 are removed from it. In other words, the inclusion of the cluster of outliers forces observations 170 and 172 to leave the subset. They re-enter only in the final two steps.

Figure 10 shows the maximum and minimum standardized one step ahead residuals. This plot immediately reveals the real structure of the data. The change in slope in the left-hand panel when the first outlier joins the subset ($m = 183$, unit 164) and in the right panel in the step prior to the inclusion of the first outlier ($m = 182$) is absolutely clear in both plots. Note that if we had used standard deletion diagnostic procedures to detect this patch of atypical observations, we should have started deleting the good units (170-172). Certainly a backwards procedure would not delete 10 observations and would fail to detect the contaminated observations. On the other hand, leaving k out when $k = 10$ is infeasible.

Table 1: Italian industrial production index with temporary level shift. Units included and removed from subset in the last 12 steps of the forward search

Step	Units included	Units removed from subset
181	140	
182	145	
183	164	
184	169	
185	166	
186	167, 168	172
187	160, 172	170
188	163, 165	172
189	162	
190	161	
191	172	
192	170	

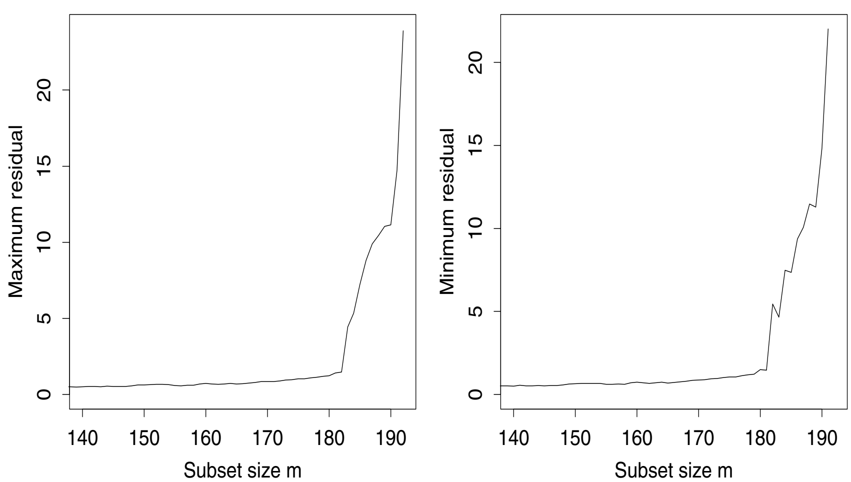

Figure 10: Italian industrial production index with temporary level shift: monitoring of the maximum squared standardized residual among the units belonging to the subset and minimum squared standardized residual among the units outside the subset

As we already pointed out in section 2, backwards procedures in time series try to detect outliers and/or structural breaks using estimated parameters based on all T observations. This implies that they will be effective in detecting atypical observations only if the final estimates of the hyperparameters are not seriously biased. Figure 11 shows the monitoring of the estimates of the parameters. The effect of the artificial temporary level shift is to pass from a smooth trend ($\sigma_\eta^2 = 0$ and $\sigma_\zeta^2 > 0$) to a non smooth trend. But, as expected, the variability in the seasonal movements as expressed by $\hat{q}_\omega$ (lower left panel of Figure 11) is virtually unaffected by the presence of these atypical observations. Finally, in the lower right hand panel we can see the monitoring of the estimate of the parameter which has been concentrated out of the likelihood (in this case $\hat{\sigma}_\epsilon^2$). Given that the forward search provides an ordering of the data according to the fitted model, we expect this curve to be increasing as the search progresses. This behaviour is similar to the monitoring of s^2 in regression (Atkinson and Riani, 2000). This panel not only shows an increasing pattern in the estimate of $\hat{\sigma}_\epsilon^2$, but also an upward jump when the first outlier joins the subset.

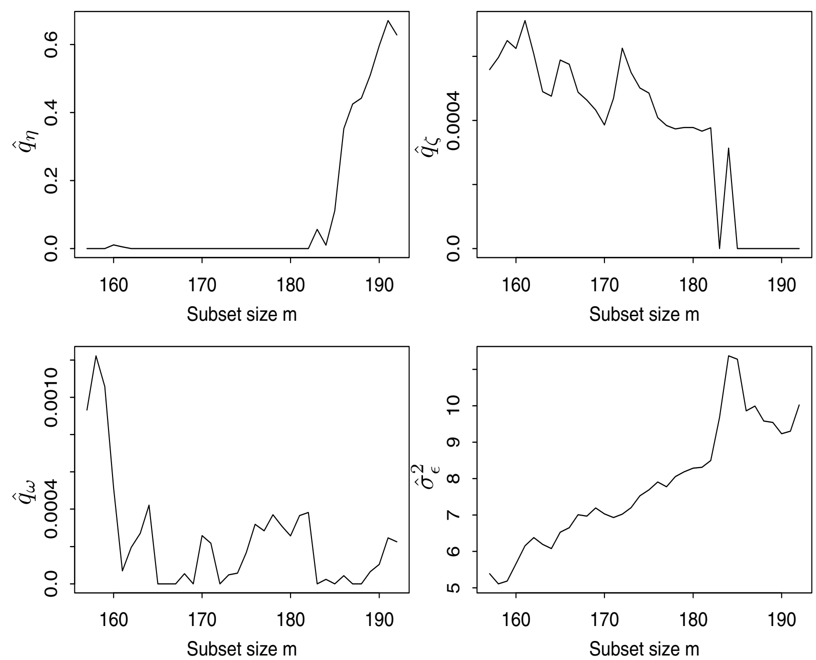

Figure 11: Italian industrial production index with temporary level shift: monitoring of the estimates of the hyperparameters

The final plot we consider is the entry plot (Figure 12). As table 1 showed, this group of 10 contaminated data enters the search in 8 consecutive steps. The white space in correspondence of units 160-169 clearly suggests that they are not in agreement with the rest of the data and that in the central part of the search they always remain outside the subset.

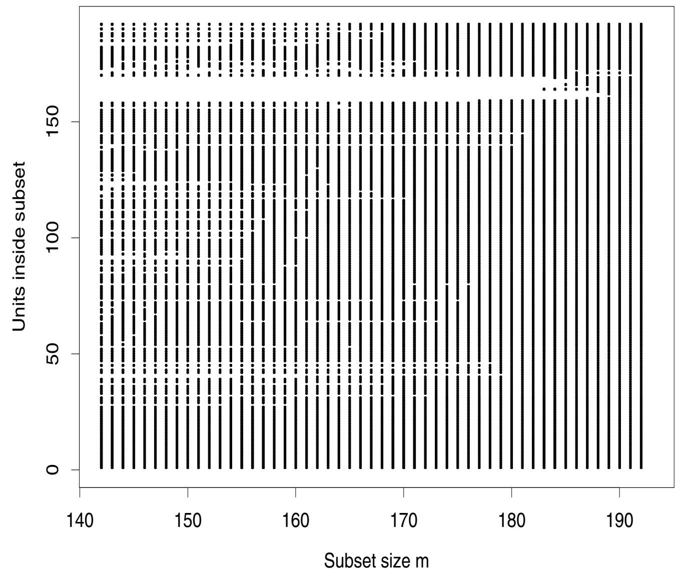

Figure 12: Italian industrial production index with temporary level shift: monitoring of units belonging to the subset in the final part of the forward search (entry plot)

4.3 American textile industrial production index

In this section we apply the suggested approach to the quarterly series of US index of industrial production for the textile sector for the period 1947.1-1996.4. Figure 13, which plots the series, shows that: 1) the dynamics at the beginning of the observation period are different from the rest of the series; 2) there is a collection of troughs and peaks of different length. The most clear one happens in the middle of the series (observations 111-115). In order to model the series Sichel (1993) and Proietti (2001) used a nonlinear cyclical component. In this section, in order to compare our results with Proietti (2003), we fit a basic structural model with linear stochastic cycle modelled as suggested in Harvey (1989).

Figure 14 shows the monitoring of one step ahead standardized prediction residuals. This plot first of all reveals that there is a series of units with unstable high residuals. Some of them are associated with the initial part of the series. Units 11 and 12 show a big downward jump when they are included in the subset in steps $m = 196$ and $m = 197$ (see Table 2). In addition, the plot shows the particular trajectory of three consecutive units 112-114. From step $m = 198$ to $m = 200$ the residual from unit 114 passes from -2.70 to 2.97 as units 112 and 113 enter.

The entry plot (Figure 15) shows patches of white spaces for the initial part of the series and strips for observations 111-115 and 140-143.

Figure 16 shows in the left panel the trajectories associated with observations 111-116 and clearly identifies the different role each observation in this stretch

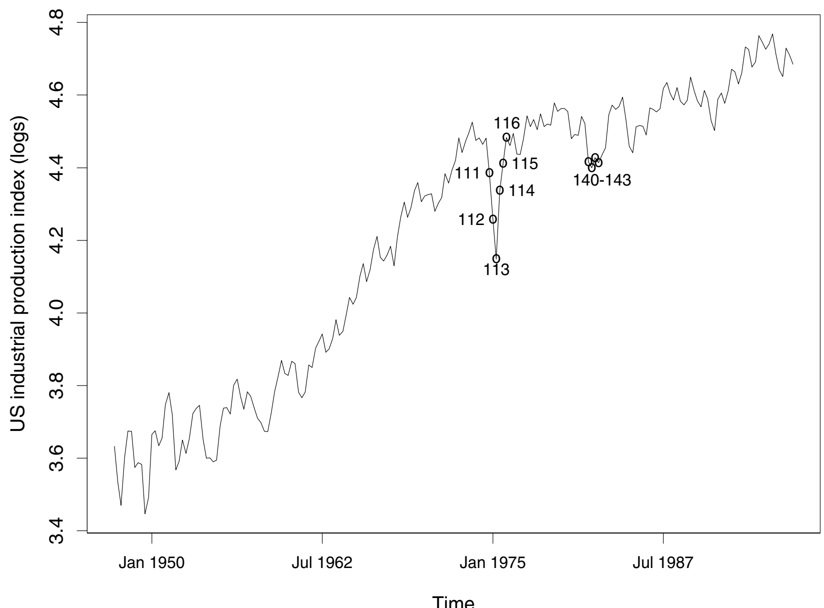

Figure 13: US industrial production textiles index: plot of the log of the series

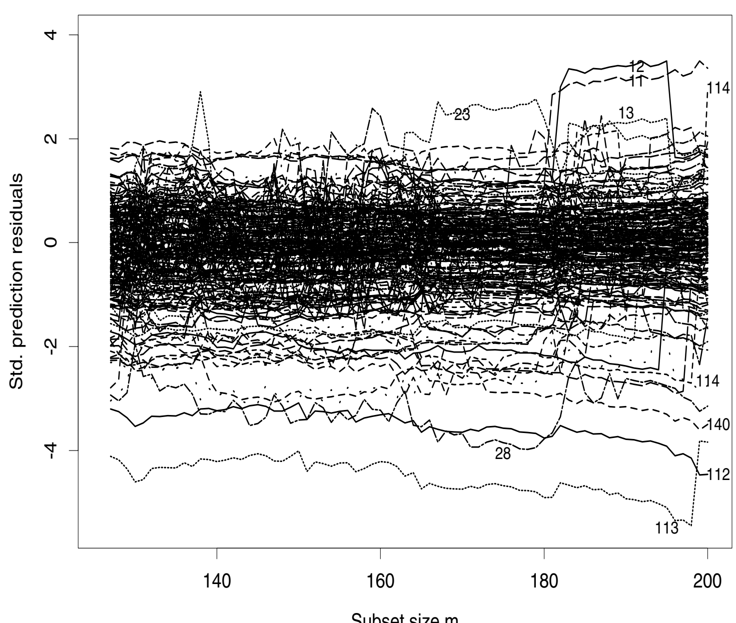

Figure 14: US industrial production textiles index: monitoring of the one step ahead standardized prediction residuals

Table 2: US industrial production textiles index: units included and removed from subset in the last 13 steps of the forward search

Step	Units included	Units removed from subset
188	14, 23	15
189	15, 29	23
190	13,23	29
191	29	
192	142	
193	114	
194	28	
195	141	
196	11	
197	12	
198	140	
199	112	
200	113	

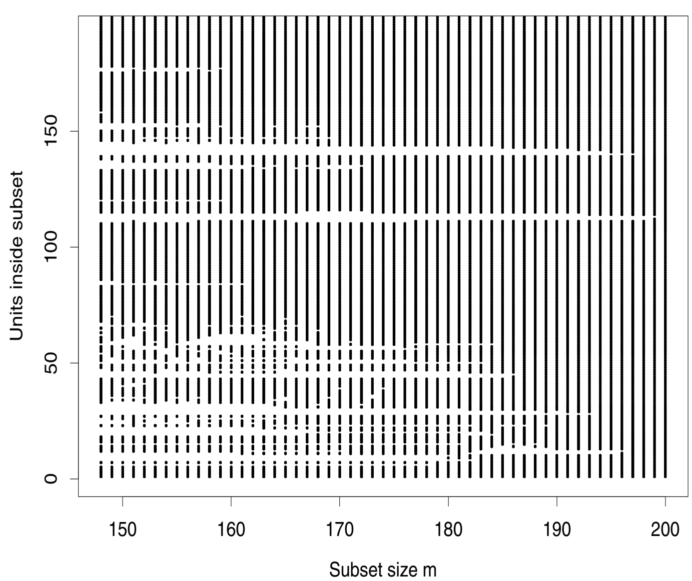

Figure 15: US industrial production textiles index: monitoring of units belonging to the subset in final part of the forward search (entry plot)

has. Unit 111 has a residual always close to -1 during all steps of the search. Units 112 and 113 have a residual which increases in absolute value as the search progresses. Unit 115 has an upward jump when $m = 193$ (inclusion of unit 114). The behaviour of the trajectory associated with unit 116 is stable and always close to zero. This plot clearly shows when the lack of fit starts and when it finishes. On the other hand, the conclusions which come from leaving out k observations are not clear. When we leave out two consecutive observations the results from Proietti (2003) show that we can only detect two atypical observations. Finally, leaving 3 out flags four consecutive outliers in the area between observations 100 and 115. The forward search, on the other hand, free from masking and smearing effects makes clear the inferential effect of each unit on the results of the model and clearly identifies where the area of lack of fit starts and when it finishes.

The trajectories associated with the other area highlighted by the entry plot (observations 140-143) are shown in the right hand panel of Figure 16. In the central part of the search, units 140-143 always have a residual around -2. On the contrary, while unit 140 always shows a stable negative residual, units 141-143 show an upward jump near the inclusion of an adjacent unit. Given that this group of units (as shown in Figure 13) forms a trough, it is clear that their presence in the estimation process lowers considerably the height of the fitted curve.

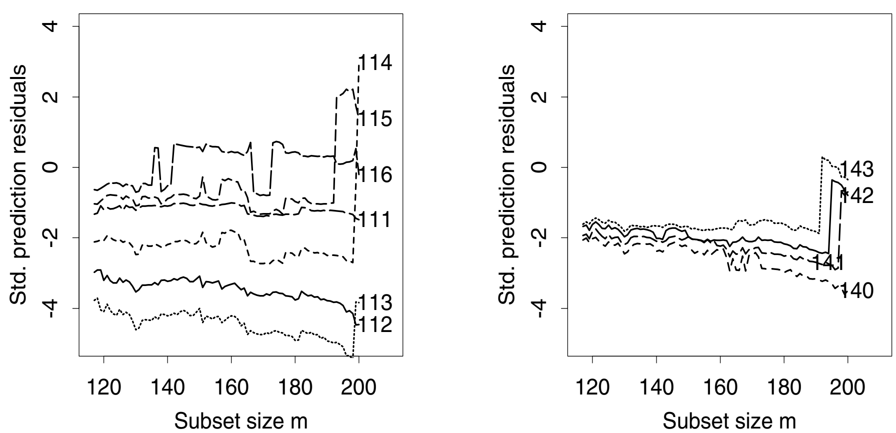

Figure 16: US industrial production textiles index: monitoring of one step ahead standardized prediction residuals for selected units

The monitoring of the hyperparameters estimates (not shown) suggests that such estimates are quite stable throughout the forward search. The frequency of the estimated cycle fluctuates around 0.5 suggesting that our estimate of the periodicity of the extracted cycle is virtually unaffected by the inclusion of the last observations. This is an example in which the diagnostics based on parameters estimated using all observations can be used for outlier detection. On the other hand, as we showed in our previous example, sometimes the estimates are seriously affected by outliers. In this last case we expect that only a procedure based on multiple deletion and reestimation of the parameters, or the application of the forward search, will be able to reveal the real structure of the data.

5 Discussion and extensions for further research

Our examples show some of the ways in which the forward search is a powerful tool for exploring the structure of time series data. Possible extensions for further research are the monitoring of smoothed auxiliary residuals (Koopman and Harvey, 1992) which have been used in the past to detect structural breaks and the results of the diagnostic tests such as heteroskedasticity or normality.

Forward plots of standardized prediction residuals, as in Figure 2, are taken directly from the output of our Gauss program. On the screen the variety of line types and colours, together with the ability to zoom, makes it possible to follow the trajectories of individual units in a way which is impossible on the printed page. Our goal is to be able to brush linked plots from the forward search, although this is still in the future.

Colophon

The author wishes to thank Anthony Atkinson, Paolo Paruolo, Sergio Zani and an anonymous referee for useful comments on a previous version of this paper. The usual disclaimer applies.
Correspondence to Marco Riani (marco.riani@unipr.it), Department of Economics - Section of Statistics and Computing, University of Parma, via Kennedy, 6, 43100 Parma, Italy.

References

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

Atkinson, A.C., Riani, M. (2001), "Regression Diagnostics for Binomial Data from the Forward Search", *The Statistician*, **50**, 63-78.

Atkinson, A.C., Riani, M. (2002), "Forward search added-variable t-tests and the effect of masked outliers on model selection", *Biometrika*, **89**, 939-946.

Atkinson, A.C., Riani, M. and Cerioli, A. (2003), *The Forward Search and Multivariate Data Analysis*, in press, New York: Springer-Verlag.

Bruce, A.M., Martin, R. D. (1989), "Leave-k-out diagnostic for time series", *Journal of the Royal Statistical Society*, **51**, 363-424.

Chen, C., Liu, L.M. (1993), "Joint estimation of model parameters and outlier effects in time series", *Journal of the American Statistical Association*, **88**, 284-297.

De Jong, P., Chu Chun Lin, S. (1994a), "Stationary and non-stationary state space models", *Journal of Time Series Analysis*, **15**, 151-166.

De Jong, P., Chu Chun Lin, S. (1994b), "Fast likelihood evaluation and prediction for nonstationary state space models", *Biometrika*, **81**, 133-142.

De Jong, P. and Penzer J. R. (1998), "Diagnosing shocks in time series", *Journal of the American Statistical Association*, **93**, 796-806.

Durbin J., Koopman S.J. (2001), *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

Gomez, V., Maravall, A. (1994), "Estimation prediction and interpolation for nonstationary time series with the Kalman filter", *Journal of Time Series Analysis*, **89**, 611-624.

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

Harvey, A.C., Pierse, R.G. (1984), "Estimating missing observations in economic time series", *Journal of the American Statistical Association*, **79**, 125-131.

Haegerty, P.J., Lumley, T. (2000), "Window subsampling of of estimating functions with application to regression models", *Journal of the American Statistical Association*, **95**, 197-211.

Hampel, F.R. Ronchetti, E.M. Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley and Sons New York.

Kaiser, R., Maravall, A. (2001), "Seasonal outliers in time series", *Estadistica, Journal of the Inter-American Statistical Association*, **53**, 97-142.

Koopman, S.J., Harvey, A. C. (1992), "Diagnostic checking of unobserved components time series models", *Journal of Business and Economic Statistics*, **10**, 377-389.

Proietti, T. (2001), "Characterizing Asymmetries in Business Cycles Using Smooth-Transition Structural Time-Series Models", *Studies in Nonlinear Dynamics & Econometrics*, **3**, 141-156.

Proietti, T. (2003), "Leave-k-out Diagnostics in State Space Models", *Journal of Time Series Analysis*, **24**, 221-236.

Riani, M. and Atkinson, A.C. (2001), "A Unified Approach to Outliers, Influence and Transformations in Discriminant Analysis", *Journal of Computational and Graphical Statistics*, **10**, 513-544.

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

Rousseeuw, P.J. and Basset, G.W. (1991), "Robustness of the p-Subset Algorithm for Regression with High Breakdown Point", in *Directions in Robust Statistics and Diagnostics, Part II*, edited by W. Stahel and S. Weisberg. The IMA Volumes in Mathematics and Its Appplications, Vol. 34, Springer Verlag, pp. 185-194.

Sichel, D.E. (1993), "Business Cycle Asymmetry: a deeper look", *Economic Inquiry*, **31**, 224-236.

Tsay, R.S. (1986), "Time series model specification in the presence of outliers", *Journal of the American Statistical Association*, **81**, 132-141.