A robust data-driven version of the Berlin Method*

Yuanhua Feng and Siegfried Heiler

University of Konstanz

Abstract

In this paper a robust data-driven procedure for decomposing seasonal time series based on a generalized Berlin Method (BV, Berliner Verfahren) as proposed by Heiler and Michels (1994) is discussed. The basic robust algorithm used here is an adaptation of the LOWESS (LOcally Weighted Scatterplot Smoothing) procedure (Cleveland, 1979). For selecting the optimal bandwidth the simple double smoothing rule (Heiler and Feng, 1999) is used. The optimal order of the local polynomial is selected with a BIC criterion. The proposed procedure is applied to the macroeconomic time series used in the recent empirical studies carried out by the German Federal Statistical Office (Speth, 1994 and Höpfner, 1998).

1 Introduction

Decomposing seasonal time series into unobservable trend-cyclical and seasonal components has a long tradition. It is a very important issue of econometrics and provides adjusted data for a prospective analysis (e.g. for a current business cycle analysis). There exists a large number of different methodical approaches and also ready-made software systems to perform this. See Heiler (1995) for a survey of methods in this field from the beginning of the 1960s. See also Eurostat (1998). Heiler (1966, 1970) developed a decomposition procedure based on local regression with polynomials and trigonometric

*This paper summarizes a part of the dissertation of Dr. Y. Feng: "Kernel- and Locally Weighted Regression – with applications to time series decomposition", Berlin, 1999, which awarded the *Förderpreis* of the *Gerhard-Fürst-Preis* of the German Federal Statistical Office, 1999. The original work was supported in part by a Doctoral-Fellowship of the State Baden-Wüttemberg, Germany and by the SFB 178, University of Konstanz, Germany.

functions as local regressors. This idea became the basis of the so-called Berlin Method (BV, Berliner Verfahren), which in its fourth version (BV 4) is being applied by the German Federal Statistical Office since 1983. Comparisons among different approaches for time series decomposition including the BV 4 with empirical macroeconomic data were carried out by the German Federal Statistical Office (Speth, 1994 and Höpfner, 1998).

The traditional idea of local regression was generalized by Stone (1977) and Cleveland (1979) to a so-called locally weighted regression (LWR), which became the most attractive nonparametric approach in recent years (see the monograph of Fan and Gijbels, 1996). This approach together with other new developments in the area of nonparametric statistics and in the area of computer science since the eighties allows us to propose an improved version of the BV. Heiler and Michels (1994) generalized the BV based on LWR by introducing a kernel as weight function. Like with other nonparametric approaches, effective use of the generalized BV requires the choice of some smoothing parameters, such as the order of the polynomial and the bandwidth. Hence, the development of some data-driven procedures for carrying out the generalized BV automatically is a crucial theoretical and practical problem. The first data-driven version of the BV was developed by Heiler and Feng (1996, 1999). In this procedure techniques to deal with outliers in the data are not considered yet.

However, the locally weighted regression estimator is susceptible to outliers among the data due to the fact that at a point t only a part of the observations is used in the local smoothing. The proposal of Cleveland (1979) is a robust approach, which works well for common nonparametric regression. In this paper we adapt at first the idea of Cleveland (1979) to the decomposition of seasonal time series with given smoothing parameters. Then we develop a robust data-driven procedure for time series decomposition using the simple double smoothing (DS) rule for the bandwidth selection (Feng and Heiler, 1999) and the BIC (Bayesian information criterion) for the choice of the polynomial degree (Feng, 1999). The proposed robust data-driven procedure is applied to the macroeconomic time series used by the recent empirical studies carried out by the German Federal Statistical Office to show its usefulness in practice.

2 The estimators

2.1 Generalized Berlin Method

The BV is based on local least squares. In BV 4 locally weighted least squares are introduced with a fixed weighting function (Nourney, 1983). This approach is generalized by Heiler and Michels (1994) based on LWR. The basic idea is as follows. Let Y_t, $t = 1, ..., n$, be an equidistant time series. Assume that (possibly after some transformation of the original data) Y_t follows an additive components model

$$Y_t = G(t) + S(t) + \epsilon_t, \quad t = 1, 2, ..., n, \tag{1}$$

where the ϵ_t are assumed to be i.i.d. random variables with $E(\epsilon_t) = 0$ and $\mathrm{var}(\epsilon_\mathrm{t}) = \sigma^2$. G is the trend-cyclical component, S is the seasonal component with seasonal period s and $m := G + S$ is the mean function.

The trend-cyclical component is assumed to have some smoothness properties, precisely, to be at least (p+1) times differentiable, so that it can be expanded in a Taylor series around a point t_0, yielding a local polynomial representation of order p

$$G(t) \doteq \sum_{j=0}^{p} \beta_{1j}(t_0)(t - t_0)^j.$$

In a similar way for the seasonal component it is assumed that it can be locally modeled by a Fourier series

$$S(t) \doteq \sum_{j=1}^{q} \left[\beta_{2j}(t_0) \cos \lambda_j (t - t_0) + \beta_{3j}(t_0) \sin \lambda_j (t - t_0)\right],$$

where $q = [s/2]$ with $[x] = \mathit{largest\ integer} \leq x$, $\lambda_1 = 2\pi/s$ is the seasonal frequency and $\lambda_j = j\lambda_1$, for $j = 2, ..., q$.

Put

$$\beta_1(t_0) = (\beta_{10}(t_0), \ldots, \beta_{1p}(t_0))^{\mathrm{T}},$$

$$
\begin{aligned}
\beta_2(t_0) &= (\beta_{21}(t_0), \beta_{31}(t_0), ..., \beta_{2q}(t_0), [\beta_{3q}(t_0)])^{\mathrm{T}}, \\
\beta(t_0) &= (\beta_1(t_0)^{\mathrm{T}}, \beta_2(t_0)^{\mathrm{T}})^{\mathrm{T}},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{x}_1(t) &= (1,\ (t-t_0),\ \dots,(t-t_0)^p)^{\mathrm{T}}, \\
\mathbf{x}_2(t) &= (\cos\lambda_1(t-t_0), \sin\lambda_1(t-t_0), ..., \\
&\quad \cos\lambda_q(t-t_0), [\sin\lambda_q(t-t_0)])^{\mathrm{T}}, \\
\mathbf{x}(t) &= (\mathbf{x}_1(t)^{\mathrm{T}}, \mathbf{x}_2(t)^{\mathrm{T}})^{\mathrm{T}}
\end{aligned}
$$

and $\mathbf{X} = (\mathbf{X_1} \vdots \mathbf{X_2})$ with the rows $\mathbf{x}(\mathbf{t})^{\mathrm{T}}$, where $\mathbf{X}$ is the $n \times (p+s)$-regressor matrix. The last terms of β_2 and $\mathbf{x}_2$ in [ ], respectively, are only necessary for odd s, for even s they have to be omitted.

Let $K(u)$, the weighting function of LWR, be a second order kernel with compact support [-1, 1], $h \in \mathbb{N}$ be a bandwidth of the observation time, $\mathbf{y} = (Y_1, \dots, Y_n)^{\mathrm{T}}$ denote the data vector and $\mathbf{K} = diag(k_t)$ be a weight matrix with

$$
k_t = \begin{cases} K(\frac{t-t_0}{h+0.5}), & t \in [t_0 - h, t_0 + h], \\ 0, & \text{otherwise.} \end{cases}
$$

Let $\mathbf{e}_j$ denote the jth $(p+1) \times 1$ unit vector, Φ_{s} denote an $(s-1) \times 1$ vector having 1 in its odd entries and 0 elsewhere.

The locally weighted least squares criterion

$$
\sum_{t=1}^{n} [Y_t - \mathbf{x}_1(t)^{\mathrm{T}}\beta_1(t_0) - \mathbf{x}_2(t)^{\mathrm{T}}\beta_2(t_0)]^2 k_t \Rightarrow \min,
$$

leads to the solutions

$$
\hat{\beta}(t_0) = (\mathbf{X}(t_0)^{\mathrm{T}}\mathbf{K}(t_0)\mathbf{X}(t_0))^{-1}\mathbf{X}(t_0)^{\mathrm{T}}\mathbf{K}(t_0)\mathbf{y} \tag{2}
$$

and

$$
\hat{m}(t_0) = (\mathbf{e}_1^{\mathrm{T}}, \Phi_{\mathrm{s}}^{\mathrm{T}})(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{y} =: \mathbf{w}^{\mathrm{T}}\mathbf{y}, \tag{3}
$$

$$
\hat{G}(t_0) = (\mathbf{e}_1^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{y} =: \mathbf{w}_1^{\mathrm{T}}\mathbf{y}, \tag{4}
$$

and

$$
\hat{S}(t_0) = (\mathbf{0}^{\mathrm{T}}, \Phi_{\mathrm{s}}^{\mathrm{T}})(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{y} =: \mathbf{w}_2^{\mathrm{T}}\mathbf{y}, \tag{5}
$$

where $\mathbf{0}$ denotes a vector of zeros of appropriate dimension.

$\hat{m}$, $\hat{G}$ and $\hat{S}$ are all linear smoothers and only observations with $t \in [t_0 - h, t_0 + h]$ obtain non-zero weights. The non-zero parts of $\mathbf{w}$, $\mathbf{w}_1$ and $\mathbf{w}_2$ will be called the weighting systems. It can be shown that $\mathbf{w}$ in (3) satisfies:

$$\begin{array}{lll}
1. & w_i(t) = 0, & \text{if } |i-t| > b, \\
2. & \sum\limits_{i=1}^{n} w_i(t)((i-t)/n)^j = \begin{cases} 1, & j = 0, \\ 0, & 1 \le j < k, \end{cases} & \\
2'. & \begin{cases} \sum\limits_{i=1}^{n} w_i(t)\cos(\lambda_j(i-t)) = 1, \\ \sum\limits_{i=1}^{n} w_i(t)\sin(\lambda_j(i-t)) = 0, \end{cases} & j = 1, \ \dots \ , \eta, \\
3. & \tilde{\mathbf{w}}'\tilde{\mathbf{K}}^{-1}\tilde{\mathbf{w}} = \min! & \text{with respect to } \tilde{\mathbf{w}},
\end{array} \tag{6}$$

where $\tilde{\mathbf{K}}$ is the non-zero part of $\mathbf{K}$ and $\tilde{\mathbf{w}}$ is the non-zero part of $\mathbf{w}$. On the other hand, weighting system satisfies (6) is the solution of (3). Properties 2 and 2' in (6) ensure that the proposed estimators are unbiased for a polynomial trend-cyclical component and for an exactly periodic seasonal component. For estimations at another instant in the central part of the time series ($t_0 \in [h+1, n-h]$) we obtain the same weighting systems (i.e. the procedure works like a moving average in the central part of the time series).

As in local polynomial fitting the νth derivative of the trend-cyclical component, $0 < \nu \le p$, can be estimated by

$$\begin{aligned}
\hat{G}^{(\nu)}(t_0) &= \nu!(\mathbf{e}_{\nu+1}^{\mathrm{T}}, \mathbf{0}^{\mathrm{T}})(\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{K}\mathbf{y} \\
&=: (\mathbf{w}^{\nu})^{\mathrm{T}}\mathbf{y}.
\end{aligned} \tag{7}$$

Finite sample and asymptotic properties of these estimators may be found in Feng (1999).

2.2 Approach at the boundary

A point $t \in [1, h] \cup [n-h+1, n]$ is called a boundary point. At such a point the observations introduced in the estimate are not symmetric around t. The quality of the estimation at a boundary point is hence usually worse than that at a point in the central part of the time series. This is the so-called boundary problem. In order to cope with this problem we distinguish a left bandwidth h_l and a right bandwidth h_r and put $h = \max(h_\mathrm{l}, h_\mathrm{r})$. It is assumed that $h_\mathrm{l} < t$, $h_\mathrm{r} \le n - t$ and $h_\mathrm{T} > p + s$, where $h_\mathrm{T} = h_\mathrm{l} + h_\mathrm{r} + 1$ is called the total bandwidth. If h_l and h_r are fixed and $t \in [h_\mathrm{l} + 1, n - h_\mathrm{r}]$, then the estimators $\hat{m}$, $\hat{G}$, $\hat{G}^{(\nu)}$ and $\hat{S}$ work as moving averages with symmetric weighting systems for $h_\mathrm{l} = h_\mathrm{r}$ or asymmetric weighting systems for $h_\mathrm{l} \neq h_\mathrm{r}$. In the proposal of Heiler and Feng (1996, 1999) both, h_T and p are allowed to change from point to point at the boundary in order to obtain optimal decomposition results. However, the estimations obtained by this procedure are sometimes not stable at the boundary. This problem is solved here by putting h_T and p fixed, i.e. just one optimal total bandwidth $\hat{h}_\mathrm{T}$ together with one optimal order of polynomial $\hat{p}$ will be selected for the whole time series. In this case the proposed estimators are like k-NN estimators. The use of k-NN estimators to solve the boundary problem was proposed by Gasser et al. (1985).

A k-NN estimator is an estimator with a fixed total bandwidth $h_\mathrm{T} = h_\mathrm{l} + h_\mathrm{r} + 1$, which is kept constant at the boundary as well as in the central part. The left bandwidth h_l and the right bandwidth h_r depend on t. Here only odd integers $h_\mathrm{T} \in [p + s + 1, n]$ will be considered as possible total bandwidth such that the weighting system in the central part is symmetric. Let $h_\mathrm{m} = (h_\mathrm{T} - 1)/2$. Then the left and right bandwidths at each point t are determined by

$$\begin{array}{lll} h_\mathrm{l} = t - 1, & h_\mathrm{r} = h_\mathrm{T} - t, & \text{if } t \le h_\mathrm{m}, \\ h_\mathrm{l} = h_\mathrm{m}, & h_\mathrm{r} = h_\mathrm{m}, & \text{if } h_\mathrm{m} < t \le n - h_\mathrm{m}, \\ h_\mathrm{l} = h_\mathrm{T} - (n - t) - 1, & h_\mathrm{r} = n - t, & \text{otherwise.} \end{array} \tag{8}$$

A nonrobust data-driven procedure for choosing h_m and p for the k-NN estimator will be introduced in the next section.

3 Nonrobust data-driven procedure

From here on only estimation of m, G and S will be considered. The optimal choice of the parameters h_T and p is the one, which minimizes a given error criterion as a distance measure between the estimate and the (unknown) underlying function. In this paper m is considered as the target function, i.e. $\hat{m}$ will be optimized with respect to m. As error criterion for selecting the bandwidth we use the mean averaged squared error (MASE)

$$M = n^{-1} \sum_{t=1}^{n} E[\hat{m}_t - m_t]^2. \tag{9}$$

It is well known that the MASE splits up into a variance part V and a bias part B, i.e. $M = V + B$. Under the assumption that ϵ_t are iid random variables the variance part of a k-NN estimator is

$$\hat{V} = \hat{\sigma}^2 n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{n} (w_i(t))^2 \,,$$

where $\hat{\sigma}^2$ is an estimator of σ^2. In this paper the bootstrap variance estimator $\hat{\sigma}^2 := \hat{\sigma}^2_{\mathrm{B}}$ as proposed by Heiler and Feng (1999) will be used. $\hat{\sigma}^2_{\mathrm{B}}$ is defined by

$$\hat{\sigma}^2_{\mathrm{B}} = n^{-1} \sum_{i=1}^{n} r_i^2 \;\;,$$

where the r_i's are the residuals of a data-driven pilot smoothing. This estimator is just the averaged residual sum of squares of a pilot estimate. A simple estimate of M was proposed by Rice (1984):

$$\tilde{R} = n^{-1} \sum_{t=1}^{n} (\hat{m}_t - y_t)^2 + (n^{-1} \sum_{t=1}^{n} 2w_t(t) - 1)\hat{\sigma}^2 \,. \tag{10}$$

This idea will be used as a pilot method for selecting the parameters. However, the pilot estimate of M in our program is actually $\hat{R} := \max(\tilde{R}, \hat{V})$, called the R-statistic, due to the fact that $M \geq V$. $\hat{R}$ depends on the couple $\{h_{\mathrm{T}}, p\}$. The optimal bandwidth selected by the R-statistic for a given p, denoted by $\hat{h}_{\mathrm{T,p}}$, is the one among all possible h_{T}'s, which minimizes $\hat{R}(p)$.

In Heiler and Feng (1996) p is also selected with the error criterion itself. In this paper we propose selecting p by means of the information criteria BIC (Schwarz, 1978 and Akaike, 1979). However, the optimal polynomial order will only be chosen from $p = 0, 1, ..., 4$, because it should not be too high. Theoretically, an odd p is more preferable (see Fan and Gijbels, 1995, 1996). But $\hat{m}$ with p even performs sometimes the best for finite sample. The original BIC was proposed for model selection in a parametric case, where BIC is a consistent criterion for model choice. In the current case we will use the following definition of BIC:

$$\text{BIC}(p) = \ln(\hat{R}(p)) + ln(n)(p+1)/n. \tag{11}$$

The term $ln(n)(p+1)/n$ is a penalty for increasing the polynomial order. The BIC depends on the couple $\{h_{\mathrm{T}}, p\}$, too. The optimal choice is the couple $\{\hat{h}_{\mathrm{T,R}}, \hat{p}\}$, which minimizes the BIC. For a given p the optimal bandwidth selected by BIC is the same as that selected by $\hat{R}(p)$ itself. But the final selected optimal couple $\{\hat{h}_{\mathrm{T,R}}, \hat{p}\}$ by BIC may be different from that directly selected by the R-statistic.

The procedure to search $\{\hat{h}_{\mathrm{T,R}}, \hat{p}\}$ for a k-NN estimator with fixed p is much simpler than that proposed in Heiler and Feng (1996). Let $h_{\max} = n$, if n is odd, or $h_{\max} = n-1$, if n is even. For $p = 0, 1, ..., 4$, let $h_{\min} = p+s+2$, if $p+s$ is odd, or $h_{\min} = p+s+3$, if $p+s$ is even. Search $\hat{h}_{\mathrm{T,p}}$ which minimizes BIC(p). Select the couple $\{\hat{h}_{\mathrm{T,R}}, \hat{p}\}$ with the smallest value of BIC from all $\{\hat{h}_{\mathrm{T,p}}, p\}$. $\{\hat{h}_{\mathrm{T,R}}, \hat{p}\}$ are then the optimal parameters selection by BIC.

The polynomial order p will only be selected once by BIC means the R-criterion. In the following we will discuss how to select a more effective bandwidth by the double smoothing rule. From now on it is assumed that we have selected a $\hat{p}$. For the DS rule one also needs a pilot estimate. As shown in Heiler and Feng (1996), the polynomial order in the pilot smoothing should be higher than that in the main smoothing. Hence, $\hat{p}_{\mathrm{p}} = \hat{p} + 2$ will be used in the pilot smoothing. The DS estimate of MASE is then defined by

$$\hat{M}_{\mathrm{D}} = \hat{V} + \hat{B}_{\mathrm{D}}, \tag{12}$$

where

$$\hat{B}_{\mathrm{D}} = n^{-1} \sum_{t=1}^{n} \{\sum_{i=1}^{n} w_i(t) \hat{m}_{\mathrm{p,i}} - \hat{m}_{\mathrm{p,t}}\}^2,$$

and where $\hat{m}_{\mathrm{p}}$ is the pilot estimate of m obtained with $p_{\mathrm{p}} = p+2$ and bandwidth $\hat{h}_{\mathrm{T,g}}$ selected by the R-statistic. The optimal bandwidth $\hat{h}_{\mathrm{T,D}}$ is the one among all possible total bandwidths, which minimizes $\hat{M}_{\mathrm{D}}$. A simple nonrobust data-driven procedure is as follows:

1. Estimate the variance with $\hat{\sigma}^2_{\mathrm{B}}$ following the proposal in Heiler and Feng (1999).

2. Select an optimal order of polynomial $\hat{p}$ following the BIC with $\hat{\sigma}^2_{\mathrm{B}}$.

3. Select an optimal total bandwidth $\hat{h}_{\mathrm{p}}$ following the R-statistic with $\hat{\sigma}^2_{\mathrm{B}}$ and $p_p = \hat{p}+2$. Calculate the pilot estimate $\hat{m}_{\mathrm{p}}$.

4. Select an optimal total bandwidth $\hat{h}_{\mathrm{T}}$ following the DS criterion with $\hat{\sigma}^2_{\mathrm{B}}$, $\hat{p}$ and $\hat{m}_{\mathrm{p}}$. Calculate all estimators one needs with $\hat{h}_{\mathrm{T}}$.

This is a simplified procedure of the proposal of Heiler and Feng (1996, 1999) due to the use of the k-NN approach and a fixed order of polynomial for all observation points of the time series.

4 Robust procedure with given parameters

It is shown that LOWESS works well in common nonparametric regression (see the examples given in Cleveland, 1979 and in Fan and Gijbels, 1996). Cleveland et al. (1990) adapted the LOWESS to time series decomposition, where seasonal fluctuations are treated in a different way. In this section we will propose a robust procedure for time series decomposition based on another adaptation of the LOWESS. Our proposal differs from the original LOWESS in two points: 1. The so-called season-dependent medians are introduced, so that the LOWESS may be easily adapted to our model. 2. A stability criterion is used, which allows the number of the robust iterations (NRI) to be decided by the data. In this section the parameters h_{T} and p are assumed to be chosen by hand. The robust data-driven version of the BV will be introduced in the next section. Detailed discussion on the properties of such a robust procedure is omitted.

The basic idea of the LOWESS is as follows. Fit $\hat{m}_0$ in the 0-th (nonrobust) iteration with weights $k_{0,i}(t) = k_i(t)$ as defined in section 2.1. Calculate the residuals $r_{j,t} = y_t - \hat{m}_{j-1}(t)$ in the j-th $(1 \leq j \leq J)$ iteration from the estimate obtained in the $j-1$-th iteration, where J is the number of robust iterations (NRI) given beforehand. Assign the robustness weight to each observation as $\beta_{j,t} = B(r_{j,t}/(6\delta_j))$, $t = 1, 2, \ldots, n$, where δ_j denotes the median of $|r_{j,t}|$ and $B(u) = (1-u^2)^2 \mathbb{I}_{[-1,1]}$ is the bisquare weighting function. Fit $\hat{m}_j$ with the weights $k_{0,i}(t)$ being replaced by $k_{j,i}(t) = \beta_{j,i} k_{0,i}(t)$. However, in the case of time series decomposition the residuals often depend on the seasonal component. Hence, a uniform median may not be suitable in this context. Therefore it will be natural to treat the residuals within each seasonal period differently. For this purpose we will use the season-dependent medians, $\delta_{j,t}$, rather than a uniform median δ_j, where $\delta_{j,t}$ is the median of $|r_{j,i}|$ for all i such that $(i-t)/s$ is an integer. This means that the residuals are at first divided into s groups. The robustness weights are then calculated in each group following the idea of Cleveland (1979). We will see that the use of season-dependent medians works well for decomposing seasonal time series.

As in Cleveland (1979) the function $B(u) = (1-u^2)^2 \mathbb{I}_{[-1,1]}$ will be used to calculate the robustness weights, which can, of cause, be replaced by another symmetric nonnegative kernel function. Assume that fixed p and h_T are used in all iterations. Then the adaptation of the LOWESS into the time series decomposition context works as follows:

1. In the 0-th iteration of this procedure decompose the time series by locally weighted regression using the weights $k_{0,i}(t) = k_i(t)$ and obtain $\hat{m}_0$.

2. In the jth iteration, let $r_{j,t} = y_t - \hat{m}_{j-1}(t)$, $t = 1, 2, \ldots, n$, denote the residuals obtained from the (j-1)th iteration. Define the robustness weights by

$$\beta_{j,t} = B\left(\frac{r_{j,t}}{6\delta_{j,t}}\right), \qquad t = 1, 2, \ldots, n,$$

   where $\delta_{j,t}$ is the season-dependent median as described above.

3. Decompose the time series with $k_i(t)$ replaced by the modified weights $k_{j,i}(t) = \beta_{j,i} k_{0,i}(t)$.

4. Repeatedly carry out steps 2 and 3 until a stability criterion is fulfilled or up to a given number of iterations.

An important question is how many robust iterations should be carried out in oder to obtain satisfactory results. It is clear that at least two robust iterations are needed, since in the first step the robustness weights are calculated from of residuals obtained by a nonrobust procedure. In the following we will propose a stability criterion such that the number NRI can be determined by the data. To simplify the description we put $\beta_{0,i} \equiv 1$ for all i in the 0-th iteration.

For given p and bandwidth h_{T} the weights $k_{j,i}(t)$ in the jth iteration depend only on $\beta_{j,i}$. Hence the estimates in the jth iteration will be close to those in the (j-1)th iteration, if $\beta_{j,i} \simeq \beta_{(j-1),i}$ for all i. Observing that $0 \leq \beta_{j,i} \leq 1$ for all j and i, we can use the averaged absolute difference (AAD) between $\beta_{j,i}$ and $\beta_{(j-1),i}$:

$$\mathrm{AAD}_j = n^{-1} \sum_{i=1}^{n} |\beta_{j,i} - \beta_{(j-1),i}|, \ j = 1, 2, ...,$$

as a measure of stability of the robust estimates in the jth iteration. For given p and h_{T} AAD_j will converge to zero, as $j \to \infty$, if the robust procedure is stable. Hence we can choose a small positive constant c_0 and stop the iterative procedure when $\mathrm{AAD}_j < c_0$. Such an estimator will be called a stable robust estimate, or simply a robust estimate, if no confusion is to be expected. The robust estimate depends strongly on c_0. If c_0 is too large, the results will not be satisfactory. If c_0 is too small, the computer time will be unnecessarily large. In this paper $c_0 = 0.0125$ will be used.

5 Data-driven robust procedure

The error criteria given in section 3 are proposed for a linear smoother under the assumption that the ϵ_t are iid. It is clear that the robust locally weighted regression estimators are nonlinear, since the robustness weights $\beta_{j,t}$ for $j > 0$ depend on the data. In this case the weighting system $\mathbf{w}$ also depends on the data. From the robust procedure described in the last section we can see that the dependence of the weighting system on the data is very complex. In this paper, however, we will simply use the R-statistic and DS criterion as approximate methods to select the order of polynomial and the bandwidth in a robust iteration. This is just an attempt to develop a robust data-driven time series decomposition procedure. There are still many open questions in this area (see the final remarks).

The extension of the bandwidth selection procedure to a robust iteration, when the nonlinearity is ignored, is straightforward. Both, the pilot bandwidth and the main bandwidth have to be reselected in each iterative step, since the optimal bandwidth in the next robust iteration may be different from that in the last one. Hence the data-driven robust procedure needs large computing time. The local regression estimates depend strongly on the bandwidth h_{T}. Besides the stability condition $\mathrm{AAD}_j < c_0$ an additional stability condition on the bandwidth will also be used. Denote the bandwidth for the main smoothing in the jth iteration as $\hat{h}_{\mathrm{T,j}}$. Assume that $j \geq 2$. Then the data-driven procedure will be stopped when $\mathrm{AAD}_j < c_0$ and $\hat{h}_{\mathrm{T,j}} = \hat{h}_{\mathrm{T,(j-1)}}$. When the procedure is not stable, it will be stopped after J iterations. In this case the best NRI has to be chosen subjectively by analyzing the detailed results at each iteration.

The proposed data-driven robust time series decomposition procedure reads as

1. For $j = 0$ obtain $\hat{p}$, the pilot bandwidth $\hat{h}_{\mathrm{p,0}}$, $\hat{h}_{\mathrm{T,0}}$ and compute $\hat{m}_0$ following the nonrobust data-driven procedure as given in section 3.

2. In the j-th, $j > 0$, iteration compute the robustness weights from $\hat{m}_{(j-1)}$ obtained in the (j-1)-th iteration.

3. Select the optimal pilot bandwidth $\hat{h}_{\mathrm{p,j}}$ following the R-statistic with the robust procedure. Calculate the robust pilot estimate $\hat{m}_{\mathrm{pj}}$.

4. Select the optimal bandwidth $\hat{h}_{\mathrm{T,j}}$ following the DS criterion with the robust procedure. Calculate the robust estimate $\hat{m}_j$.

5. For $j \geq 2$ put $j_0 := j$ and go to step 6, if the stability conditions are satisfied or $j_0 = J$. Otherwise, put $j := j + 1$ and go back to step 2.

6. Calculate other estimates with the selected parameters and the robust procedure.

The selected parameters of such a procedure is the polynomial order $\hat{p}$ and a sequence of bandwidths $\{\hat{h}_{\mathrm{T,0}}, \hat{h}_{\mathrm{T,1}}, ..., \hat{h}_{\mathrm{T,j_0}}\}$. The estimates at the end of this procedure depend not only on $\hat{h}_{\mathrm{T,j_0}}$ but on the whole sequence $\{\hat{h}_{\mathrm{T,0}}, \hat{h}_{\mathrm{T,1}}, ..., \hat{h}_{\mathrm{T,j_0}}\}$. If we let the robust procedure proposed in the last section, run $j_0 + 1$ times with $\hat{p}$ and corresponding bandwidth, $\hat{h}_{\mathrm{T,j}}$, in each iteration, we will obtained the same estimates.

6 Applications

In this section the proposed procedure will be applied to the seven examples used by the recent empirical studies carried out by the German Federal Statistical Office (Speth, 1994 and Höpfner, 1998). These are observations in the old States in Germany from January 1976 to June 1987 for the following time series:

1. Unemployment (UNEMP),
2. Production index of production industries (PIPIN),
3. Production index of automobile industry (PIAUT),
4. Production index of manufacture of tobacco (PITOB),
5. Production index of chemical industry (PICHE),
6. Index of orders received in engineering industry (OIENG),
7. Index of orders received in building construction (OIBUI).

Here $c_0 = 0.0125$ and $J = 20$ are used for decomposing these time series. The estimated parameters $\hat{\sigma}^2_{\mathrm{B}}$, $\hat{p}$, j_0 and $\hat{h}_{\mathrm{T}} = \hat{h}_{\mathrm{T},j_0}$ are listed in Table 1. The bandwidths selected in other iterations are omitted. From Table 1 we see that the selected parameters for these time series are quite different. This shows that a model with given parameters can not be suitable for all data sets in practice. Hence a data-driven procedure is required. The selected $\hat{p}$ and $\hat{h}_{\mathrm{T}}$ depend on the variance σ^2 of a time series and also on the structure of the trend-cyclical component. The larger σ^2 (relatively), the larger $\hat{h}_{\mathrm{T}}$. On the other hand, the more complex G is, the smaller is $\hat{h}_{\mathrm{T}}$. Furthermore, $\hat{p}$ and $\hat{h}_{\mathrm{T}}$ depend also mutually on each other. The higher $\hat{p}$, the larger $\hat{h}_T$, and vice versa. For all of the examples we have obtained stables robust decomposition results, except for the time series OIBIU, for which we have $j_0 = J$. For this time series the optimal bandwidth switches between $\hat{h}_T = 49$ and $\hat{h}_T^* = 37$. If $J = 19$ or $J = 21$ were used, we would have obtained $\hat{h}_T^*$ as the final result.

Table 1: $\hat{\sigma}_B^2$, $\hat{p}$, j_0 and $\hat{h}_T$ selected for all examples

Time Series	$\hat{\sigma}_B^2$	$\hat{p}$	j_0	$\hat{h}_T$
UNEMP	470.79	4	10	55
PIPIN	3.767	2	7	55
PIAUT	57.20	0	4	63
PITOB	15.55	0	8	43
PICHE	2.734	2	7	37
OIENG	28.01	1	4	41
OIBUI	5.709	2	20	49

Decomposition results for the three time series PIAUT, PITOB and OIENG, which seem to be affected by outliers at some points, are shown in figures 1 through 3. Each of the figure gives the original data together with the nonrobust as well as the stable robust estimates of G (upper), the nonrobust estimate of S (middle) as well the stable robust estimate of S (lower), respectively. We see, in all of these examples both, the estimates of G and S are improved by the robust procedure at instants, where there seem to be some outliers. Figure 2 shows that (see the observations around $t = 80$) the effect of outliers, which are difficult to make out at the first glance, can be corrected by means of the season-dependent medians. And the corrected seasonal component looks much better than the nonrobust one. For more examples see Feng (1999).

7 Final remarks

In this paper a robust data-driven version of the BV is proposed. Examples from the practice show that the proposed procedure works well. However, there are some open questions to be resolved in future. For instance, some error criteria, which are more suitable for selecting optimal parameters by the robust time series decomposition procedure as those proposed in section 4, should be developed. The variance estimator $\hat{\sigma}^2_{\mathrm{B}}$ used in this paper is nonrobust. A robust variance estimator for seasonal time series should be developed. If the variance of ϵ_t depends strongly on the seasonal component, then both, the error criterion and the variance estimator should be adapted to this fact. The finite sample properties and asymptotic properties of the parameters selected by the robust data-driven procedure should be investigated. Moreover, the error process variables ϵ_t are generally not independent. In this case it is more difficult to solve the above mentioned questions. Finally, note that the selected parameters are just optimal for the estimate of m. They are neither optimal for $\hat{G}$ nor for $\hat{S}$. The development of procedures which yield optimal parameters for $\hat{G}$ or $\hat{S}$ alone, are also very important.

8 Acknowledgements

This paper was supported in part by the Center of Finance and Econometrics at the University of Konstanz, Germany. We would like to thank the colleagues in the Department IIA, German Federal Statistical Office for providing us with the data.

References

[1] Akaike, H. (1979). A bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika*, **66**, 1979, 237–242.

[2] Cleveland, R.B., Cleveland, W.S., McRae, I.E. and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on LOWESS (with discussion). *J. Official Statistics*, **6**, 3–73.

[3] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. J. Amer. Statist. Assoc., **74**, 829–836.

[4] Eurostat (1998). *Seasonal Adjustment Methods* – A Comparison. Luxembourg

[5] Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *J. Roy. Statist. Soc. Ser. B*, **57** 371–394.

[6] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and its Applications.* Chipman and Hall, London.

[7] Feng, Y. (1999). *Kernel- and Locally Weighted Regression* – with applications to time series decomposition. Verlag für Wissenschaften und Forschung, Berlin.

[8] Feng, Y. and Heiler, S. (1999). Selecting bandwidth for nonparametric regression based on double smoothing rule. Preprint, University of Konstanz.

[9] Gasser, Th., Müller, H.G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B*, **47**, 238–252.

[10] Heiler, S. (1960). *Analyse der Struktur Wirtschaftlicher Prozesse durch Zerlegung von Zeitreihen.* Dissertation, University of Tübingen.

[11] Heiler, S. (1970). Theoretische Grundlagen des 'Berliner Verfahrens'. In Wetzel, W. (ed.): *Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse*, Sonderheft **1** zum Allg. Statistischen Archiv, 67–93.

[12] Heiler, S. (1995). Zur Glättung saisonaler Zeitreihen. In Rinne, H., Rüger, B. and Strecker, H. (eds.): *Grundlagen der Statistik und ihre Anwendungen*, Festschrift für Kurt Weichselberger, Physika-Verlag, Heidelberg, 128–148.

[13] Heiler, S. and Feng, Y. (1996). Datengesteuerte Zerlegung saisonaler Zeitreihen. *ifo Studien*, **3/1996**, 41–73.

[14] Heiler, S. and Feng, Y. (1999). Data-driven decomposition of seasonal time series. To appear in *J. of Statistical Planning and Inference.*

[15] Heiler, S. and Michels, P. (1994). *Deskriptive und Explorative Datenanalyse.* Oldenbourg, München.

[16] Höpfner, B. (1998). Ein empirischer Vergleich neuerer Verfahren zur Saisonbereinigung und Komponentenzerlegung. *Wirtschaft und Statistik*, **12/1998**, 949–959.

[17] Nourney, M. (1983). Umstellung der Zeitreihenanalyse. *Wirtschaft und Statistik*, **11/1983**, 841–852.

[18] Rice, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215–1230.

[19] Speth, H.-Th. (1994). Vergleich von Verfahren zur Komponentenzerlegung von Zeitreihen. *Wirtschaft und Statistik*, **2/1994**, 98–108.

[20] Stone, C.J. (1977). Consistent nonparametric regression (with discussion). Ann. Statist., **5**, 595–620.

[21] Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461–464.

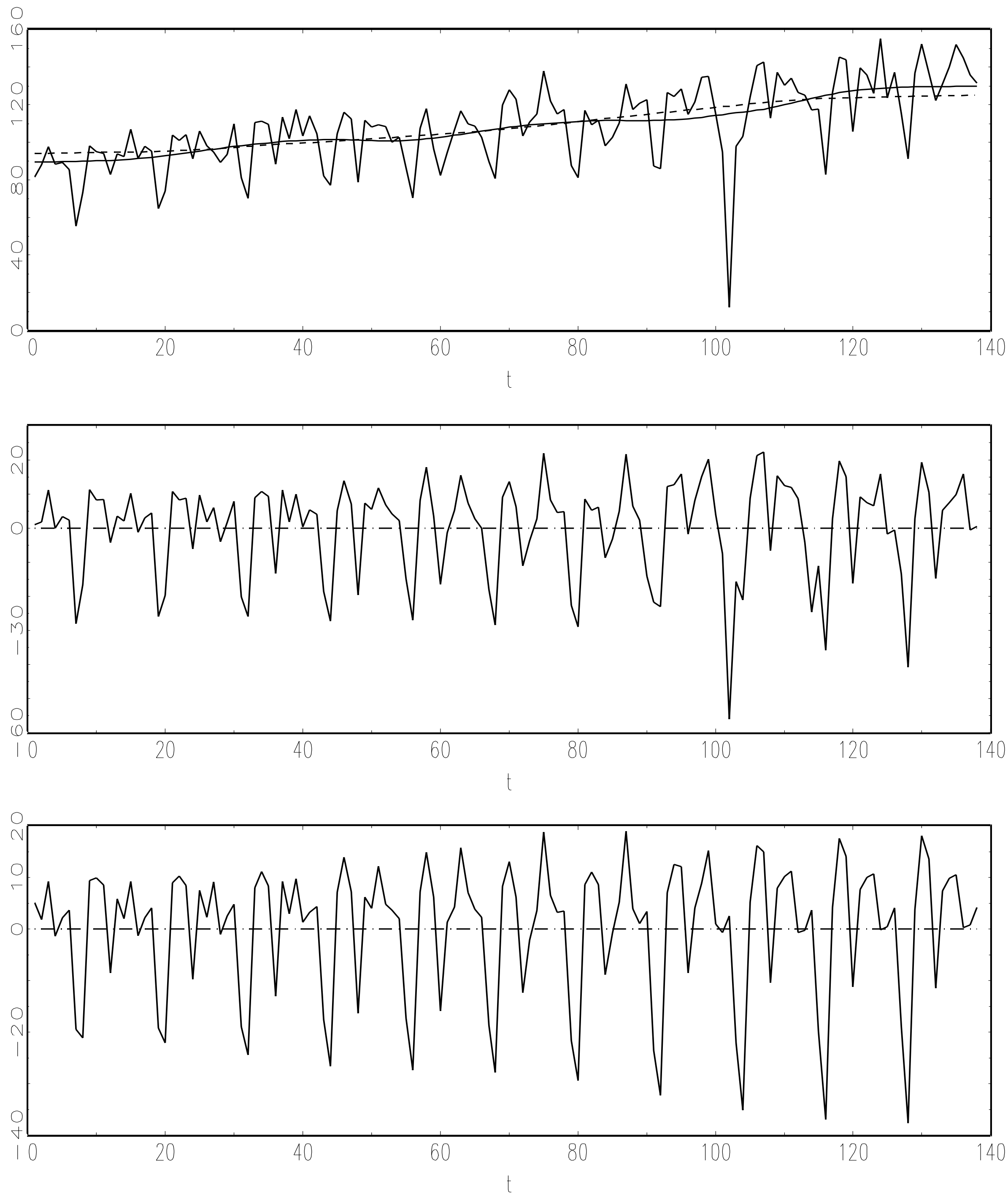

Figure 1: Nonrobust and stable robust decomposition results for the time series PIAUT. The figures show the original data together with the nonrobust (solid line) as well as the stable robust (dashes) estimates of G (upper), the nonrobust estimate of S (middle) as well the stable robust estimate of S (lower), respectively.

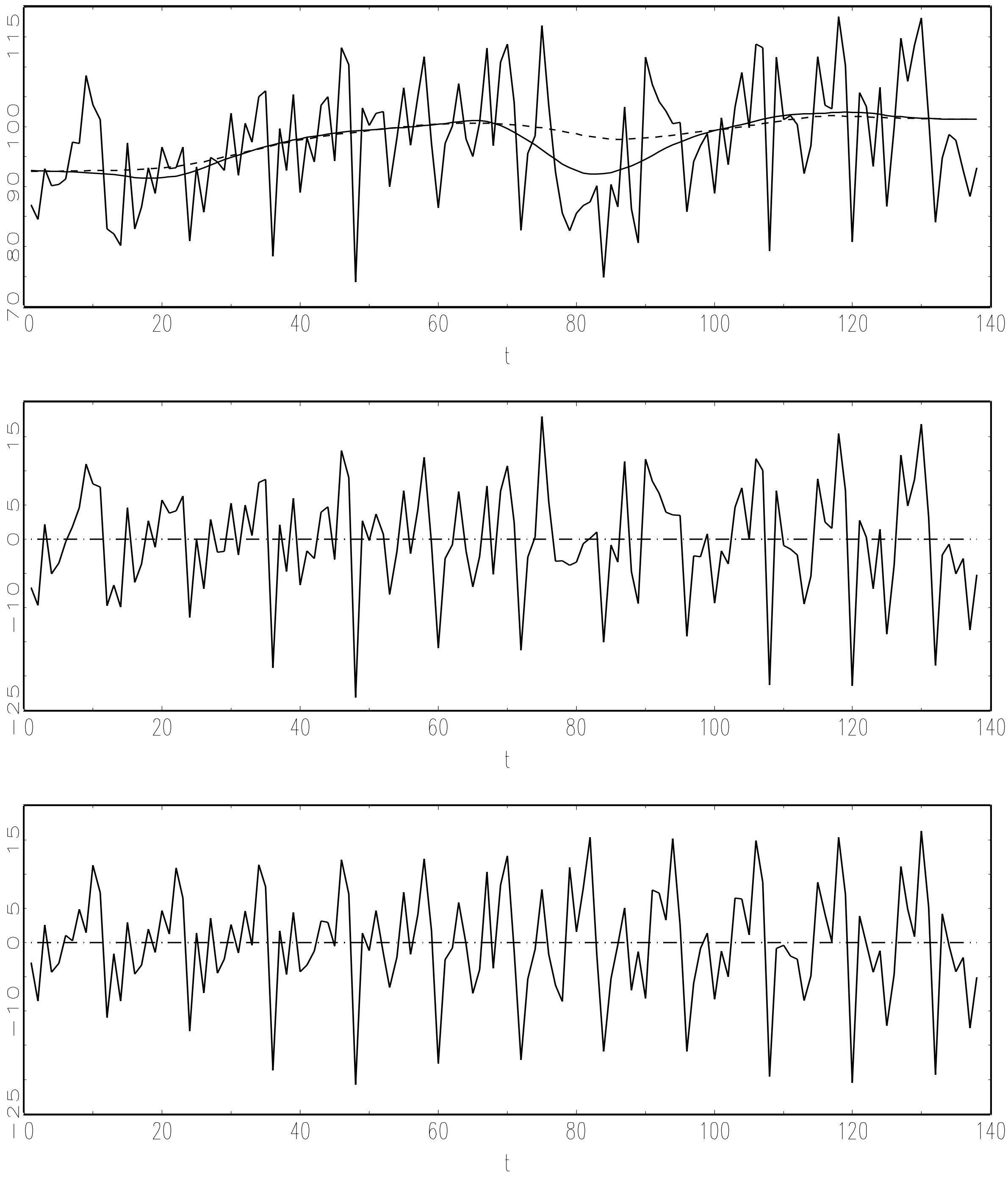

Figure 2: The same results as given in figure 1 but for the time series PITOB.

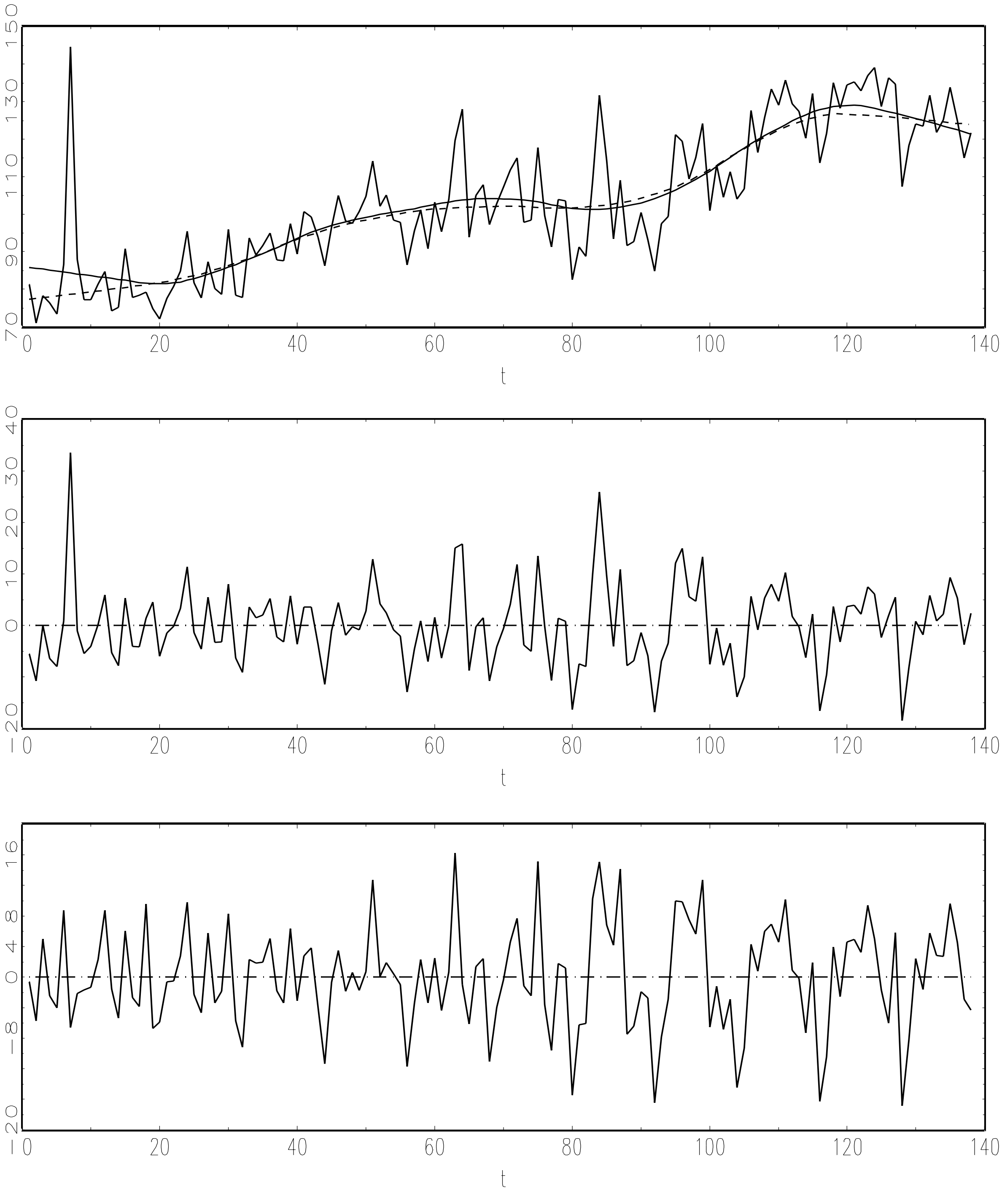

Figure 3: The same results as given in figure 1 but for the time series OIENG.