An Iterative Plug-In Algorithm for Nonparametric Modelling of Seasonal Time Series

Yuanhua Feng

University of Konstanz

Abstract

This paper focuses on developing a new data-driven procedure for decomposing seasonal time series based on local regression. Formula of the asymptotic optimal bandwidth h_{A} in the current context is given. Methods for estimating the unknowns in h_{A} are investigated. A data-driven algorithm for decomposing seasonal time series is proposed based on the iterative plug-in idea introduced by Gasser et al. (1991). Asymptotic behaviour of this algorithm is investigated. Some computational aspects are discussed in detail. Practical performance of the proposed algorithm is illustrated by simulated and data examples. The results here also provide some insights into the iterative plug-in idea.

Keywords: Time series decomposition; Local regression; Iterative plug-in; Bandwidth selection

1 Introduction

Decomposing seasonal time series into unobserved components is an important issue of statistics. This question arises, if e.g. we want to analyze monthly data or to build models using seasonally adjusted data. Here, the equidistant additive time series model

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

will be used to perform this, where $x_t = (t - 0.5)/n$, ϵ_t are iid random variables with $E(\epsilon_t) = 0$ and $\mathrm{var}(\epsilon_t) = \sigma^2$, g is a smooth trend-cyclical function, S is a slowly changing seasonal component with seasonal period s (in terms of t). Denote by $m = g + S$ the mean function. In the following model (1) will be treated as a standard nonparametric regression with an additional (deterministic) seasonal component. A traditional nonparametric approach for estimating g, S and m based on (unweighted) local regression with polynomials and trigonometric functions as local regressors was proposed by Heiler (1966, 1970). This became the basis of the so-called Berlin Method, which in its fourth version is being used by the German Federal Statistical Office since 1983. The approach used in the following is a generalized version of the Berlin Method proposed by Heiler and Michels (1994) based on locally weighted regression (Cleveland, 1979) by introducing a common kernel weight function into the original methodology.

A crucial problem for time series decomposition based on local regression is the selection of the bandwidth. Some double-smoothing procedures to perform this are proposed by Heiler and Feng (1996, 2000), Feng (1999) and Feng and Heiler (2000). The aim of the current paper is to propose a new algorithm for selecting the bandwidth under model (1). The iterative plug-in idea introduced by Gasser et al. (1991) with some minor improvements proposed by Beran and Feng (2002a, b) is adapted to the current context. To our knowledge this is the first plug-in bandwidth selector for decomposing seasonal time series. Moreover, we also provide some insights into the iterative plug-in idea. Asymptotic behaviour of the proposed algorithm is investigated. Some computational aspects of this algorithm are discussed in detail. Simulated and data examples show that this algorithm works well in practice.

The paper is organized as follows. The estimators and some of their asymptotic properties that are needed in the subsequent sections, are given in Section 2. Methods for estimating the unknown terms in the asymptotically optimal bandwidth are discussed in Section 3. The algorithm is proposed in Section 4 together with discussion on its asymptotic behaviour and on some computational aspects. Simulated and data examples in Section 5 illustrate the practical usefulness of this proposal. Section 6 contains some final remarks. Proofs of the results are put in the appendix.

2 The local regression approach

2.1 The estimators

Assume that g is at least (p+1) times continuously differentiable, so that it can be expanded in a Taylor series around a point x_t. Similarly, S can be locally modelled by a Fourier series. Let $\lambda_1 = 2\pi/s$ be the seasonal frequency and $\lambda_j = j\lambda_1$, for $j = 2, ..., q$, where $q = [s/2]$ with $[\cdot]$ denoting the integer part. Let $K(u)$ be a second order kernel function with compact support $[-1,1]$. Let h denote the bandwidth. The locally weighted regression estimators of g, S and m at x_t are obtained by solving the least square problem

$$
\begin{aligned}
Q &= \sum_{i=1}^{n}\{Y_t - \sum_{j=0}^{p}\beta_{1j}(x_i-x_t)^j \\
&\quad - \sum_{j=1}^{q}(\beta_{2j}\cos\lambda_j(i-t)+[\beta_{3j}\sin\lambda_j(i-t)])\}K\left(\frac{x_i-x_t}{h}\right) \Rightarrow \min\,. \qquad (2)
\end{aligned}
$$

The solutions of (2) are $\hat{g}(x_t)=\hat{\beta}_{10}$, $\hat{S}(x_t)=\sum_{j=1}^{q}\hat{\beta}_{2j}$ and $\hat{m}(x_t)=\hat{g}(x_t)+\hat{S}(x_t)$, where the coefficients and their estimations are defined locally and hence depend on x_t.

Let

$$
\mathbf{X}_1 = \begin{pmatrix} 1 & x_1-x_t & \cdots & (x_1-x_t)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n-x_t & \cdots & (x_n-x_t)^p \end{pmatrix}
$$

and

$$
\mathbf{X}_2 = \begin{pmatrix} \cos\lambda_1(1-t) & \sin\lambda_1(1-t) & \cdots & \cos\lambda_q(1-t) & [\sin\lambda_q(1-t)] \\ \vdots & \vdots & \ddots & \vdots & [\vdots] \\ \cos\lambda_1(n-t) & \sin\lambda_1(n-t) & \cdots & \cos\lambda_q(n-t) & [\sin\lambda_q(n-t)] \end{pmatrix}.
$$

Then $\mathbf{X}=(\mathbf{X}_1\vdots\mathbf{X}_2)$ is the $[n\times(p+s)]$-design matrix. The entries in (2) and $\mathbf{X}_2$ marked by [ ] only apply to odd s, for even s they have to be omitted due to $\lambda_q=\pi$. Let $\mathbf{y}=(y_1,\ldots,y_n)'$ be the vector of observations and $\mathbf{K}$ denote a diagonal matrix with

$$
k_i = K(\frac{x_i-x_t}{h}).
$$

Furthermore, denote the j-th $(p+1)\times 1$ unit vector by $\mathbf{e}_j$ and let Φ_s be an $(s-1)\times 1$ vector having 1 in its odd entries and 0 elsewhere. Then we have

$$
\hat{m}(x_t) = (\mathbf{e}_1', \Phi_s')(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y} =: \mathbf{w}'\mathbf{y}, \qquad (3)
$$

$$
\hat{g}(x_t) = (\mathbf{e}_1', \mathbf{0}')(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y} =: \mathbf{w}_1'\mathbf{y}, \qquad (4)
$$

and

$$
\hat{S}(x_t) = (\mathbf{0}', \Phi_s')(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y} =: \mathbf{w}_2'\mathbf{y}, \qquad (5)
$$

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

The vectors $\mathbf{w}=(w_1,\ldots,w_n)'$, $\mathbf{w}_1=(w_{11},\ldots,w_{1n})'$ and $\mathbf{w}_2=(w_{21},\ldots,w_{2n})'$ will be called weighting systems of $\hat{m}$, $\hat{g}$ and $\hat{S}$, for which we have $\mathbf{w}=\mathbf{w}_1+\mathbf{w}_2$, $\sum w_i=\sum w_{1i}=1$ and $\sum w_{2i}=0$. The local regression approach makes $\hat{m}$, $\hat{g}$ and $\hat{S}$ exactly unbiased, if g is a polynomial of order no larger than p and S is exactly periodic with period s.

2.2 Asymptotic properties

To develop a plug-in bandwidth selector we have to discuss the asymptotic behaviour of $\hat{g}$, $\hat{S}$ and $\hat{m}$. From here on it is assumed that p is odd so that $\hat{g}$ has automatic boundary correction. Put $k = p + 1$ and assume that

A1. $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

A2. g is at least k times continuously differentiable.

A3. S is exactly periodic with period s.

A1 and A2 are the same as in nonparametric regression without seasonality. A3 is a parametric assumption on the seasonal component, which is made for simplicity and can easily be relaxed to a general slowly changing seasonal component. Under the assumption A1 it can be showed that $\hat{g}$ is asymptotically equivalent to some kernel estimator (see Feng, 1999). This means that the same asymptotic results in local polynomial fitting hold for $\hat{g}$ under model (1). In the following denote by $K_p(u)$ the equivalent kernel for estimating g, which is of order k.

To deal with $\hat{S}$, we will introduce a kernel estimator of S. Let

$$Q_s(i) = \begin{cases} (s-1), & \text{if } (i-t)/s \text{ is an integer,} \\ -1, & \text{otherwise}, \end{cases} \tag{6}$$

and

$$\breve{w}_{2i} = (nh)^{-1} Q_s(i) K\left(\frac{x_i - x_t}{h}\right). \tag{7}$$

A kernel estimator of S is defined by

$$\breve{S}(x_t) = \sum_{i=1}^{n} \breve{w}_{2i} y_i =: \breve{\mathbf{w}}_2' \mathbf{y}. \tag{8}$$

Note that $\{\breve{w}_{2i}\}$ are asymptotically periodic with the same period s. Suppose that corresponding boundary correction is done for $\breve{S}$, then it can be shown that, under A1, $\hat{S}$ and $\breve{S}$ are asymptotically equivalent, too (see Feng, 1999).

As an error criterion for bandwidth selection we use the mean averaged squared error (MASE). Define $R(K) = \int_{-1}^{1} K^2(u)du$. Let B denote the bias of an estimator. We have

Lemma 1 *Assume that A1 to A3 hold, then*

1. *the asymptotic bias of* $\hat{m}$ *is*

$$B[\hat{m}(x_t)] \doteq B[\hat{g}(x_t)] \doteq \frac{1}{(k!)} \left\{ \left[\int u^k K_p(u) du \right] g^{(k)}(x_t) \right\} h^k, \tag{9}$$

2. the asymptotic variance of $\hat{m}$ *is*

$$var(\hat{m}(x_t)) = (nh)^{-1}\sigma^2\{R(K_p) + (s-1)R(K)\}\{1 + O[(nh)^{-1}]\} \tag{10}$$

3. and the MASE *of* $\hat{m}$ *is*

$$\begin{aligned} \mathrm{MASE}(\hat{m}) &:= \frac{1}{n}\sum_{t=1}^{n}[E(\hat{m}(x_t)) - m(x_t)]^2 \\ &\doteq \frac{\sigma^2}{nh}\{R(K_p) + (s-1)R(K)\} \\ &\quad + \frac{1}{(k!)^2}\left\{\int\{g^{(k)}(x)\}^2dx\left[\int u^kK(u)du\right]^2\right\}h^{2k}. \end{aligned} \tag{11}$$

A sketched proof of Lemma 1 is given in the appendix, where it is shown in particular that: 1. $\hat{g}$ and $\hat{S}$ are asymptotically uncorrelated and 2. the bias in $\hat{S}$ is negligible compared to that in $\hat{g}$. The asymptotically optimal bandwidth, which minimizes the dominate part of the MASE is given by

$$h_{\mathrm{A}} = \left(\frac{(k!)^2}{2k}\frac{\sigma^2\{R(K_p) + (s-1)R(K)\}}{\int\{g^{(k)}(x)\}^2dx\{\int u^kK_p(u)du\}^2}\right)^{1/(2k+1)} n^{-1/(2k+1)}, \tag{12}$$

where it is assumed that $I = \int\{g^{(k)}(x)\}^2dx > 0$. The change in h_{A} due to the additional term S is just a constant. For $s = 1$ the above formulae reduce to the results in nonparametric regression as given e.g. in Müller (1988), Ruppert and Wand (1994) and Fan and Gijbels (1996).

3 Estimating the unknown parameters

3.1 Estimation of the variance

In order to develop a plug-in bandwidth selector based on (12), the unknown parameters as σ^2 and I have to be estimated. It is well known that the variance in nonparametric regression can be estimated by difference-based methods (see e.g. Rice, 1984, Gasser et al., 1986 and Hall et al., 1990). Heiler and Feng (1996) adapted this idea to model (1) and proposed some seasonal-difference-based variance estimators. Here a sequence $D_{ms} = \{d_j|\ j = 0, 1, \ldots, m\}$ is called a *seasonal difference sequence*, if

$$\sum_{j=0}^{m} d_j = 0, \qquad \sum_{j=0}^{m} d_j^2 = 1, \quad m = 1, 2, \ldots \tag{13}$$

and

$$S_i = \sum_{j=0}^{m} d_j \delta_{ij} = 0, \;\; i = 0, 1, ..., s-1, \tag{14}$$

where

$$\delta_{ij} = \begin{cases} 1, & \text{if } (j-i)/s \text{ is an integer,} \\ 0, & \text{otherwise.} \end{cases}$$

A seasonal-difference-based variance estimator is then defined by

$$\hat{\sigma}^2_{\mathrm{D}} = (n-m)^{-1} \sum_{i=1}^{n-m} (\sum_{j=0}^{m} d_j Y_{i+j})^2. \tag{15}$$

Following Hall et al. (1990) it can be shown that under A2 and A3 $\hat{\sigma}^2_{\mathrm{D}}$ is root n consistent. In this paper the following seasonal difference sequence

$$D_{m,s} = \frac{1}{12}\{-1, 2, -1, \underbrace{0, ..., 0,}_{s-3} 1, -2, 1\}$$

defined for $s \geq 3$ will be used to estimate σ^2, where $m = s + 2$.

3.2 Estimation of I

Similar to local polynomial fitting the k-th derivative of g can be estimated with a local polynomial of order p_{I} and a bandwidth h_{I} with $p_{\mathrm{I}} > k$ and $p_{\mathrm{I}} - k$ odd. And we set $l = p_{\mathrm{I}} + 1$. A simple choice is $p_{\mathrm{I}} = k + 1$ with $l = k + 2$. Let now (2) be defined with p being replaced by p_{I}. Let $\mathbf{K}$, $\mathbf{y}$ and $\mathbf{e}_j$ are the same as defined in Section 2. Let $\mathbf{X}$ be defined similarly as before. Then $\hat{g}^{(k)} = k!\hat{\beta}_k$ estimates $g^{(k)}$, which is given by

$$\hat{g}^{(k)}(t) = k!(\mathbf{e}'_{k+1}, \mathbf{0}')(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y} =: (\mathbf{w}^k)'\mathbf{y}. \tag{16}$$

where $\mathbf{0}$ is the same as in (4) and $\mathbf{w}^k = (w_1^k, ..., w_n^k)'$ is the weighting system of $\hat{g}^{(k)}$. Then I may be estimated by

$$\hat{I}[g^{(k)}(x; h_{\mathrm{I}})] = n^{-1} \sum_{i=1}^{n} \{\hat{g}^{(k)}(x_i; h_{\mathrm{I}})\}^2. \tag{17}$$

In the following some results on $\hat{I}$, which are important for the development of a plug-in bandwidth selector, will be given without proof, since here we are only interested in the magnitude orders. These orders are the same in the current context and in models without seasonality. Such results in nonparametric regression may be found in Ruppert et al. (1995) and Herrmann and Gasser (1994). Related results in nonparametric regression with short memory, long memory and antipersistence are given in Beran and Feng (2002a).

Assume now that

A$'$1. $h \to 0$ and $nh^{2k+1} \to \infty$ as $n \to \infty$.

A$'$2. g is at least l times continuously differentiable.

Under the assumptions A$'$1, A$'$2 and A3 we have

$$B(\hat{I}) \doteq O(h_{\mathrm{I}}^{(l-k)}) + O[(nh_{\mathrm{I}})^{-1}h_{\mathrm{I}}^{-(2k)}] \tag{18}$$

and

$$\mathrm{var}(\hat{I}) \doteq O(n^{-1}) + O(n^{-2}h_{\mathrm{I}}^{-4k-1}). \tag{19}$$

A$'$1 implies that h_{I} is of a larger order than h_{A}, i.e. $(h_{\mathrm{I}})^{-1} = o[(h_{\mathrm{A}})^{-1}]$, which ensures that $\hat{g}^{(k)}$ and hence $\hat{I}$ is at least consistent.

The following remarks show that how h_{I} should be chosen.

Remark 1. The largest order h_{I} should take is $O(n^{-1/(4k+2)}) = O[(h_{\mathrm{A}})^{1/2}]$. Under this choice the second term on the right hand side of (18) and the standard deviation of $\hat{I}$ achieve the fastest root n convergence rate at the same time. An h_{I} of a larger order will increase the bias without improving the variance (in terms of the magnitude order).

Remark 2. The optimal bandwidth for estimating $g^{(k)}$ itself is of order $O(n^{-1/(2l+1)})$. This order lies between the two given in Remarks 1 and 3. The choice $h_{\mathrm{I}} = O(n^{-1/(2l+1)})$ is hence also reasonable.

Remark 3. Observe that the MSE (mean squared error) of $\hat{I}$ is dominated by the squared bias part. By balancing the orders of the two terms on the right hand side of (18) we obtain $h_{\mathrm{I}} = O(n^{-1/(k+l+1)})$, which may be considered to be the (asymptotically) optimal choice of h_{I}. Such a choice is of a smaller order than the two given in Remarks 1 and 2.

4 The main proposal

4.1 The basic algorithm

From here on only $p = 1$ and 3 with $k = 2$ and 4 will be considered. Following the iterative plug-in idea of Gasser et al. (1991), $\hat{I}_j$, the estimate of I in the j-th iteration, is calculated with a bandwidth $h_{\mathrm{I},j}$, which is obtained from h_{j-1}, the bandwidth for estimating m in the (j-1)-th iteration, by means of a inflation method. Here a inflation method is a function $h_{\mathrm{I},j} = f(h_{j-1})$ such that $(h_{\mathrm{I},j})^{-1} = o[(h_{j-1})^{-1}]$. This means that $h_{\mathrm{I},j}$ will be of a larger order than h_{A}, if h_{j-1} is at least of order $O(h_{\mathrm{A}})$. Now A$'$1 is satisfied so that $\hat{I}$ and $\hat{h}$ will be consistent in the j-th iteration. In the following two inflation methods will be described.

The original inflation function, called a multiplied inflation method (MIM) is $h_{\mathrm{I},j} = f(h_{j-1}) = ch_{j-1}n^{\alpha}$ introduced by Gasser et al. (1991) with some $\alpha > 0$, called a inflation factor. This idea is discussed in detail by Herrmann and Gasser (1994). There are some unknowns in the function f as c, α and a starting bandwidth h_0, which have to be chosen beforehand. Theoretically, the rate of convergence of $\hat{h}$ does not depend on c and h_0. Here we will simply set $c = 1$. The choice of h_0 will be discussed in Section 4.3. Let $l = k + 2$. Following Remarks 1 to 3, we obtain three reasonable choices of α for the MIM respectively.

1. $\alpha_1 = 1/(4k + 2)$ so that the variance term of $\hat{I}$ is minimized,

2. $\alpha_2 = 4/[(2k + 1)(2k + 5)]$ so that $\hat{g}^{(k)}$ is optimized and

3. $\alpha_3 = 2/[(2k + 1)(2k + 3)]$ so that the MSE of $\hat{I}$ is minimized,

when convergence is reached, where $\alpha_1 > \alpha_2 > \alpha_3$ and α_3 is the optimal choice of α.

It is well known that the required number of iterations by the MIM is very large, especially for $k > 2$. For example, if $k = 4$, it is $5k+1 = 21$ for α_1 and $(k+1)(2k+1) = 45$ for α_3 (see Herrmann and Gasser, 1994). The required number of iterations will be even larger, if the errors have long memory (see Ray and Tsay, 1997). Beran (1999) introduced another inflation method $h_{\mathrm{I},j} = f(h_{j-1}) = ch_{j-1}^{\beta}$, called an exponential inflation method (EIM), to reduce the number of iterations. It is easy to show that, in order to inflate h_{A} to a given order, the required number of iterations by the EIM is much smaller than by the MIM (see Beran and Feng, 2002a for some examples). In the following the EIM with $c = 1$ will be used. The choices of β corresponding to the α_1, α_2 and α_3 are:

1. $\beta_1 = 1/2$,

2. $\beta_2 = (2k + 1)/(2k + 5)$ and

3. $\beta_3 = (2k+1)/(2k+3)$,

where $\beta_1 < \beta_2 < \beta_3$ and β_3 is the optimal choice of β (see Beran and Feng, 2002a, b).

In the following we will propose a basic iterative plug-in algorithm for selecting bandwidth in time series decomposition, which is defined for $k = 2$ and $k = 4$ separately.

i) Start with a possible bandwidth h_0;

ii) For $j = 1, 2, \ldots$ set $h_{\mathrm{I},j} = h_{j-1}^{\beta}$ with $\beta = \beta_3 = 5/7$ for $k = 2$ and $\beta = \beta_2 = 9/13$ for $k = 4$. Calculate

$$h_j = \left(\frac{(k!)^2}{2k} \frac{\hat{\sigma}^2 \left\{ R(K_p) + (s-1)R(K) \right\}}{\int \{\hat{g}^{(k)}(x; h_{\mathrm{I},j})\}^2 dx \{\int u^k K_p(u) du\}^2} \right)^{1/(2k+1)} n^{-1/(2k+1)}, \tag{20}$$

iii) Increase j by 1 and repeat Step *ii)* until convergence is reached at some j^0 and set $\hat{h} = h_{j^0}$.

A closely related proposal for bandwidth selection in nonparametric regression with short memory, long memory and antipersistence is proposed by Beran and Feng (2002a). Other iterative plug-in procedures may be found in Gasser et al. (1991), Herrmann et al. (1992) and Ray and Tsay (1997).

Theoretically, β_3 is the asymptotically optimal choice of β. Our experience show that, for $k = 2$, this choice works well in practice. Hence we choose $\beta_3 = 5/7$ for $k = 2$. However, $\beta_3 = 9/11$ for $k = 4$ is too close to one and for small samples the bandwidth could not be inflated correctly. For $k = 4$ it is hence proposed to use the slightly stronger inflation factor β_2. Now, the variance of $\hat{h}$ with $k = 2$ and $k = 4$ is almost of the same order and $\hat{h}$ is hence in both cases stable (see Theorem 1 in the next subsection). The most stable inflation factor $\beta_1 = 1/2$ by the EIM is too strong and does not work well for small samples.

4.2 Asymptotic behaviour

The iterative plug-in algorithm is motivated by fixed point search. Here the procedure is started with a bandwidth h_0 and stopped, if a convergent output (a fixed point) is arrived. The inflation process behind an iterative plug-in algorithm is described by the following lemma according to the relationship between h_0 and h_{A}.

Lemma 2 *Under assumptions* $\mathrm{A}'2$ *and* A3*, an iterative plug-in algorithm processes as follows:*

Case 1. Start with an $h_0 = o_p(h_\mathrm{A})$*, then*

Step 1. $h_j = O_p(h_{\mathrm{I},j})$*, if* $h_{\mathrm{I},j} = o_p(h_\mathrm{A})$*;*

Step 2. $h_j = O_p(h_\mathrm{A})$*, if* $h_{\mathrm{I},j} = O_p(h_\mathrm{A})$*;*

Step 3. $h_j = h_\mathrm{A}[1 + o_p(1)]$*, if* $h_\mathrm{A} = o_p(h_{\mathrm{I},j})$*.*

Case 2. Start with an h_0 *such that* $(h_0)^{-1} = o_p[(h_\mathrm{A})^{-1}]$*, then*

Step 1′. $h_j = O_p(h_\mathrm{A})$*, if* $h_{\mathrm{I},j} = O_p(1)$*.*

Step 2′. The same as Step 3 in case 1.

The proof of Lemma 2 is given in the appendix. Related results may be found in Beran and Feng (2002a) (see also the description in Herrmann and Gasser, 1994, p. 8). Note in particular that $\mathrm{A}'1$ does not apply to Lemma 2.

Remark 4. Case 1 in Lemma 2 shows that, by starting with a small bandwidth, h_{j-1} will be inflated in the j-th iteration, if $h_{j-1} = o_p(h_\mathrm{A})$. This will be repeatedly carried out until $h_{j'} = O_p(h_\mathrm{A})$ is achieved in the j'-th iteration. And $h_{j'+1}$ in the next iteration will be a consistent bandwidth selector. Some further iterations are required to improve the finite sample property of $\hat{h}$.

Remark 5. Case 2 in Lemma 2 shows how such an algorithm works, if a starting bandwidth h_0, which is at least of order $O_p(h_\mathrm{A})$, is uaed. On one hand, if $h_0 = o_p(1)$, then h_1 is already consistent, since $\mathrm{A}'1$ is satisfied. In this case Step $1'$ will not appear. On the other hand, if $h_0 = O_p(1)$, then $h_1 = O_p(h_\mathrm{A})$, which is already of the correct order but not yet consistent. Now h_2 will be consistent. Again, some further iterations are needed to reduce the influence of h_0.

The following theorem hold for the algorithm proposed in Section 4.1.

Theorem 1 *Under the assumptions of Lemma 2 we have*

i) For $k = 2$ *with* $\beta_3 = 5/7$

$$\hat{h} = h_\mathrm{A} \left\{ 1 + O(n^{-2/7}) + O_p(n^{-5/14}) \right\}, \tag{21}$$

ii) For $k = 4$ with $\beta_2 = 9/13$

$$\hat{h} = h_{\mathrm{A}} \left\{ 1 + O(n^{-2/13}) + O_p(n^{-9/26}) \right\}. \tag{22}$$

A sketched proof of Theorem 1 is given in the appendix.

Remark 6. Let h_{M} denote the optimal bandwidth, which minimizes the MASE. Theorem 1 also holds, if h_{A} on the right hand sides of (21) and (22) is replaced by h_{M}. This is due to the fact that $|h_{\mathrm{M}} - h_{\mathrm{A}}|/h_{\mathrm{M}} = O(h_{\mathrm{M}}^2)$ (see Beran et al., 2000 and Beran and Feng, 2002b), which is of orders $O(n^{-2/5})$ for $k = 2$ and $O(n^{-2/9})$ for $k = 4$ and is hence negligible.

4.3 Computational aspects

This subsection deals with some computational aspects as the decision of j^0, the choice of h_0 and so on. A more practical procedure will be proposed at the end of this subsection.

The estimators in Section 2.1 are defined with a fixed bandwidth h. In this case the number of observations used at x_t decreases when x_t moves from the interior to the boundary. To solve this problem the k-NN idea as proposed by Gasser et al. (1985) will be used. For a given h we define a left bandwidth h_l and a right one h_r so that $h_l = h_r = h$ in the interior, $h_l = x_t$ at a left boundary point and $h_r = 1 - x_t$ at a right boundary point. h_r (rep. h_l) at a boundary point is decided by $h_l + h_r = 2h$. The estimates at a boundary point are calculated similarly but with h in (2) being replaced by $\max(h_l, h_r)$.

In our program only bandwidths $h \in [h_{\min}, h_{\max}]$ with $h_{\min} = s/n$ and $h_{\max} = 0.5 - 1/n$ will be considered, which includes practically all reasonable possibilities for h. Furthermore, two bandwidths h and h' will be considered to be the same, if $|h - h'| < 1/n$, because a difference of such an order is for any bandwidth selector negligible. In the program the actually used bandwidth is an integer $b = [nh + 0.5]$, which is the bandwidth w.r.t. the observation time t. Let $b_{\mathrm{I},j} = [nh_{\mathrm{I},j} + 0.5]$. Then we obtain a natural criterion for stopping the computing procedure, i.e. the procedure will be stopped, if $b_{\mathrm{I},j^0} = b_{\mathrm{I},j^0-1}$ in the j^0-th iteration. This implies $\hat{I}_j^0 = \hat{I}_{j^0-1}$ and $\hat{h} = h_{j^0} = h_{j^0-1}$. Further iterations are not necessary. Note that, even the j^0-th iteration is just a repetition of the (j^0-1)-th.

In the following the choice of h_0 will be considered. In most cases h_0 does not play any role. However, in some cases, when the finite sample MASE has more than one local minimums or when the MASE changes very slowly around its minimum, then $\hat{h}$ may depend on h_0 in some way. To explain this we will introduce some concepts. A bandwidth h_f is called a fixed point (of the procedure proposed in Section 4.1), if $\hat{h} = h_f$, when the procedure is started with $h_0 = h_f$ itself. A fixed point h_f is called left stable, if for all $h_0 \leq h_f$ in a neighbourhood of h_f we have $\hat{h} = h_f$. A fixed point h_f is called right stable, if for all $h_0 \geq h_f$ in a neighbourhood of h_f we have $\hat{h} = h_f$. A fixed point h_f is called stable, if it is both left and right stable. A fixed point is called unstable, if it is only achievable by starting with itself. An interval of bandwidths $[h_f^l, h_f^r]$ is called an interval of fixed points, if h_f^l is a left stable fixed point, h_f^r is a right stable fixed point and all points between them are unstable fixed points. Denote by $\hat{h}^l$ the bandwidth selected with $h_0^1 = h_{\min}$ and $\hat{h}^r$ the bandwidth selected with $h_0^2 = h_{\max}$. Then $\hat{h}^l$ is a left stable fixed point, if $\hat{h}^l > h_{\min}$ and $\hat{h}^r$ is a right stable fixed point, if $\hat{h}^r < h_{\max}$.

When the finite sample MASE has only one minimum, then there exits a unique stable fixed point or a unique interval of fixed points. In the first case we will obtain the same selected bandwidth $\hat{h}$ by starting with any h_0. In the second case we have $\hat{h} = h^l$ for all $h_0 \leq h^l$, $\hat{h} = h^r$ for all $h_0 \geq h^r$ and $\hat{h} = h_0$ for $h^l < h_0 < h^r$. Now all bandwidths in $[h^l, h^r]$ are reasonable to be used as the optimal bandwidth, since now the change of the MASE over $[h_f^l, h_f^r]$ is negligible. In this case we also say that the result is *unique* and set $\hat{h} := (\hat{h}^l + \hat{h}^r)/2$. In the following the words *a stable fixed point* also means sometimes an interval of stable points. In the case when the finite MASE has more than one local minimums, then we may obtain different $\hat{h}$ by starting with different h_0. Now, there may also be some unstable fixed points corresponding to a local maximum between two local minimums. If this is the case, we should find out all possible stable fixed points and then select one of them as the optimal bandwidth by analyzing the smoothing results.

An S-Plus function called DeSeaTS (Decomposing Seasonal Time Series) is developed based on the following quasi-data-driven procedure.

1. Carry out the algorithm in Section 5.1 twice with $h_0^1 = h_{\min}$ and $h_0^2 = h_{\max}$, respectively.
2. Calculate the decomposition results automatically, if $\hat{h}$ is unique.
3. Show detailed information about all stable fixed points, when $\hat{h}$ is not unique.

If 3 occurs, further subjective analysis is required.

For choosing p, we propose to carry out the above procedure with $p = 1$ and $p = 3$ respectively. If the smoothing results with $p = 1$ and $p = 3$ are both satisfactory, we can choose either $p = 1$ or $p = 3$. However, it is more preferable to use $p = 3$, since now the selected bandwidth is in general slightly larger, which does not increase the bias of $\hat{g}$ but will improve $\hat{S}$. Sometime one p is more reasonable than the other, now the reasonable one should be chosen (see the examples given in the next section). An objective criterion for choosing p is not given here, because we do not have an estimate of the MASE at the end of the procedure.

5 Practical performance

In the following, the practical performance of the proposed procedure will be investigated by simulated and data examples. Here the bisquare kernel is used as weight function.

5.1 Simulated examples

At first some simulated examples will be analyzed. Here the trend function

$$g = 2\sin(2(x - 0.5)\pi) + 2x + 4\exp(-100(x - 0.5)^2) + 6$$

is used, where $x \in [0, 1]$. Figure 1a to b show two simulated time series of length $n = 200$, called Sim1 and Sim2, generated with iid $N(0, 1)$ errors, where the exactly periodic seasonal component $S_1 = \{1.5, -1.2, -0.8, 0.5\}$ with $s = 4$ is used. The selected bandwidths $\hat{h}^l$ and $\hat{h}^r$ with $p = 1$ and $p = 3$ respectively, are given in Table 1 together with the corresponding j^0's, the value $d := n(\hat{h}^r - \hat{h}^l)$ and the answer to the question, if $\hat{h}$ is unique. We see that $\hat{h}^l$ and $\hat{h}^r$ are exactly the same in all cases. This is always the case for many

Table 1: $\hat{h}^l$, $\hat{h}^r$ and other parameters for the simulated examples

Time	$p = 1$						$p = 3$					
Series	$\hat{h}^l$	j^0	$\hat{h}^l$	j^0	d	uniq.	$\hat{h}^l$	j^0	$\hat{h}^r$	j^0	d	uniq.
Sim1&3	0.139	7	0.139	5	0	Yes	0.142	7	0.142	6	0	Yes
Sim2&4	0.141	6	0.141	6	0	Yes	0.163	11	0.163	6	0	Yes

other simulated examples we have done.

Another two simulated time series, called Sim3 and Sim4, with the same trend and corresponding errors as by Sim1 and Sim2 respectively, but another seasonal component $S_2 = \{0.3, -0.5, 0.9, -0.7\}$ are shown in Figures 1c to d. With these examples it is shown that the change of the seasonal component does not change the selected bandwidth. The selected bandwidths, other parameters and even the detailed information in each iteration for Sim3 and Sim4 are exactly the same as for Sim1 and Sim2 respectively.

Data-driven decomposition results with $p = 3$ for the simulated time series are shown in Figures 1a to d, where the data are given in the upper part of each figure (solid line) together with the underlying trend function (dotted line) and the estimated trend (dashed line). The estimated seasonal component is plotted in the lower part of each figure. From Figure 1 we see that the proposed procedure works well. $\hat{g}$ in all cases is satisfactory. Although at first glance we are not sure if these time series are seasonal or not, especially by Sim3 and Sim4, the seasonal component is well discovered from the data. It is clear that S_1 is more easily to estimate than S_2, since $\text{var}(\hat{S}_1) = \text{var}(\hat{S}_2)$ but the absolute values of S_1 are related larger than those of S_2. Note that the selected bandwidth under model (1) for a simulated time series with given errors will be the same, even if the seasonal component is set to zero. Now $\hat{S}$ is fully due to the noises in the data.

5.2 Data examples

The following data examples are chosen to illustrate the practical performance of the proposal in common cases and to show some details.

1. Time Series "CAPE" – Time series of the quarterly final consumption expenditure in Australia (total private, millions of dollars, 1989/90 prices) from September 1959 to June 1995 with $n = 144$. Source: Australian Bureau of Statistics.

2. Time Series "Strom" – The monthly time series of produced electricity in Germany from 1955 to 1979 with $n = 300$. Source: Schlittgen and Streitberg (1994, p. 82).

3. Time Series "IFOR" – The monthly time series of the indices of the foreign orders received in Germany from 1978 to 1994 (1985 = 100) with $n = 204$. Source: IFO-Institute for Economic Research in Munich.

4. Time Series "Hsales" – Monthly sales of new one-family houses sold in the USA from

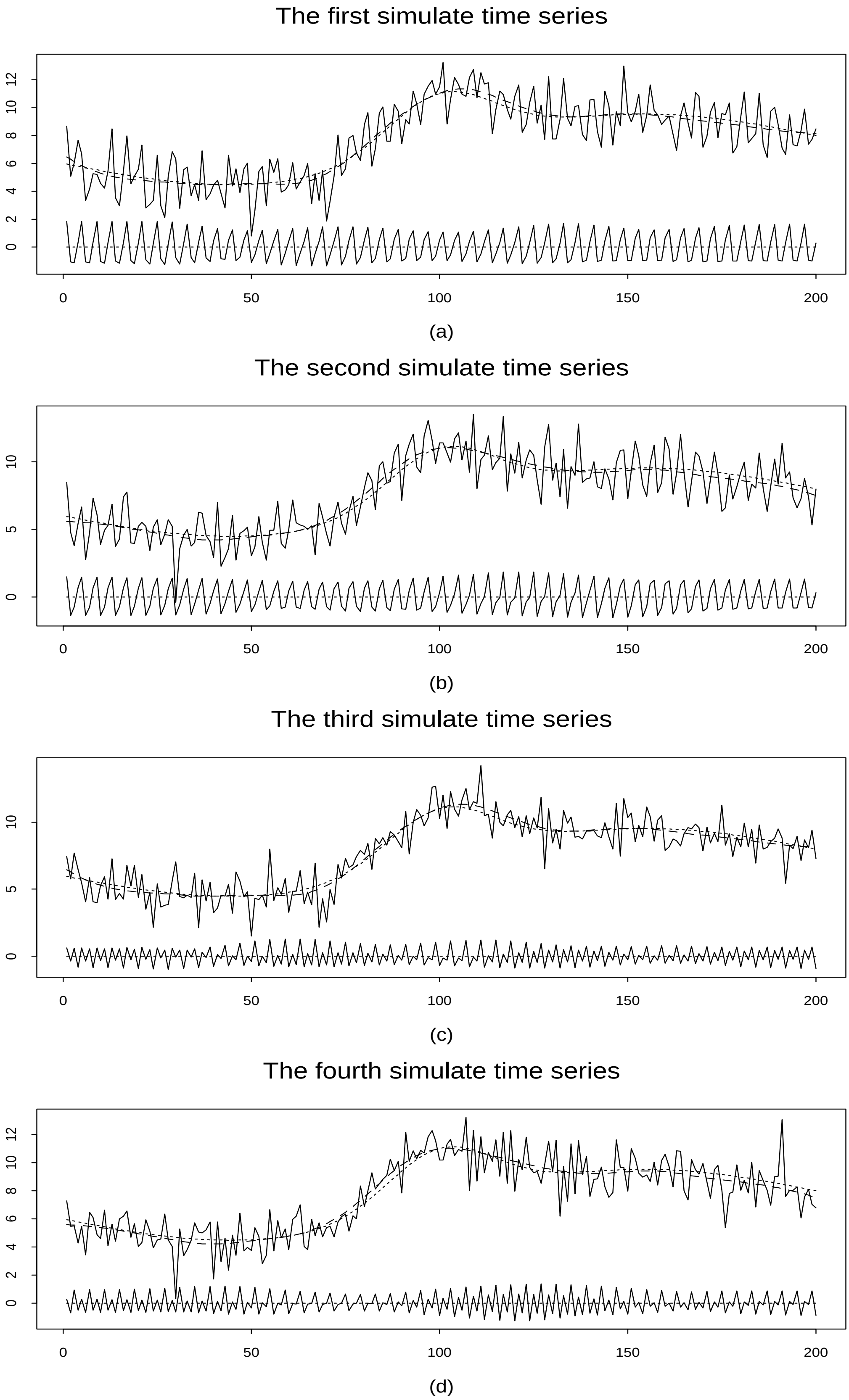


Figure 1: Optimal decomposition results for the simulated time series. Upper: the data together with the underlying trend (points) and the estimated trend (dashes). Below: the estimated seasonal component.

January 1973 to November 1995 with $n = 275$. Source: Makridakis, Wheelwright and Hyndman (1998).

All of these time series are analyzed with $p = 1$ and $p = 3$ respectively. Table 2 shows the same parameters for these data examples as those given in Table 1. From Table 2 we see that the results in most of the cases are unique. For the series Hsales with $p = 3$ we obtained an interval of fixed points, $[0.094, 0.105]$. As mentioned before, we will consider such a result to be *unique* and now $\hat{h} = (0.105 + 0.094)/2 = 0.10$ will be used. Two unusual cases are: Firstly, the selected bandwidths for the series IFOR with $p = 1$ are not unique; Secondly, although the selected bandwidth for the series Strom with $p = 3$ is unique, which is however much smaller than that selected for the same series with $p = 1$. These two cases will be discussed in the next subsection in detail.

Table 2: $\hat{h}^l$, $\hat{h}^r$ and other parameters for the data examples

Time	$p = 1$						$p = 3$					
Series	$\hat{h}^l$	j^0	$\hat{h}^r$	j^0	d	uniq.	$\hat{h}^l$	j^0	$\hat{h}^r$	j^0	d	uniq.
CAPE	0.084	7	0.086	6	0.288	Yes	0.089	6	0.089	8	0	Yes
Strom	0.160	7	0.160	7	0	Yes	0.101	7	0.102	13	0.300	Yes
IFOR	0.113	6	0.262	3	30.40	No	0.140	7	0.141	6	0.204	Yes
Hsales	0.066	4	0.067	8	0.257	Yes	0.094	7	0.105	4	3.025	Int

As explained in the last section, the use of $p = 3$ is more preferable. But for the series Strom $p = 1$ should be used (see the next subsection for reason). Hence $p = 1$ for the series Strom and $p = 3$ for the other is chosen. Data-driven decomposition results for these examples are shown in Figures 2a trough d, where corresponding location changes are introduced for the seasonal component so that the figures look more clear. We see that the results given in Figure 2 look quite well. This shows the practical usefulness of the proposed procedure. Note that the selected bandwidths for the examples given in Figure 2a to d are quite different, which adapt automatically to the structure of the data. The largest is $\hat{h} = 0.16$ by the series Strom. This is not surprising, because the trend in this time series can almost be modelled by a parametric model (see Schlittgen and Streitberg, 1994). Although the trend in the time series CAPE is also regular, the selected bandwidth $\hat{h} = 0.089$ is however the smallest one, since $s = 4$ for this time series but for the other $s = 12$. Tables 1 and 2 also show that j^0 changes from case to case.

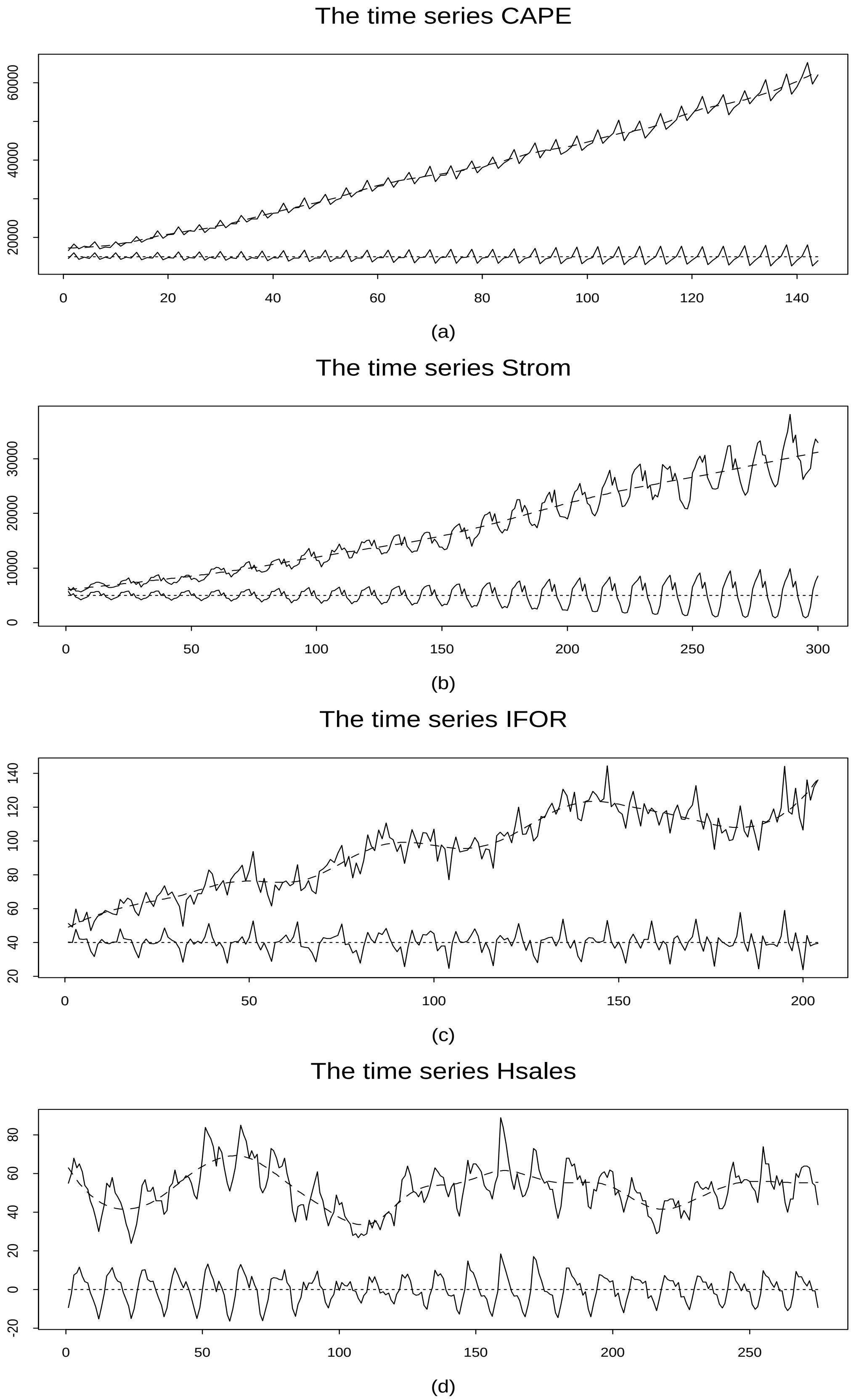


Figure 2: Optimal decomposition results for the data examples. Upper: the data together with the estimated trend (dashes). Below: the estimated seasonal component.

5.3 Some details

Following Lemma 2 we have $\hat{h}^l \leq h_A \leq \hat{h}^r$ in probability. From Tables 1 and 2 we see that this are all true for the examples. Lemma 2 also ensures that, in probability, h_j is nondecreasing in j by starting with h_0^1 and h_j is nonincreasing in j by starting with h_0^2. The detailed search processes with starting bandwidths h_0^1 and h_0^2 respectively are shown in Figure 3, where the results are for the time series Strom with $p = 1$ (solid line) and CAPE with $p = 3$ (dashed line). These two examples are chosen so that the iterative plug-in algorithm can be well understood. From Figure 3 we can find an interesting phenomenon, i.e. although the selected bandwidth for Strom with $p = 1$ is much larger than that for CAPE with $p = 3$, h_1 with h_0^2 in the second case was even slightly larger than that in the first case. However, after some iterations both of them arrived at the corresponding fixed points.

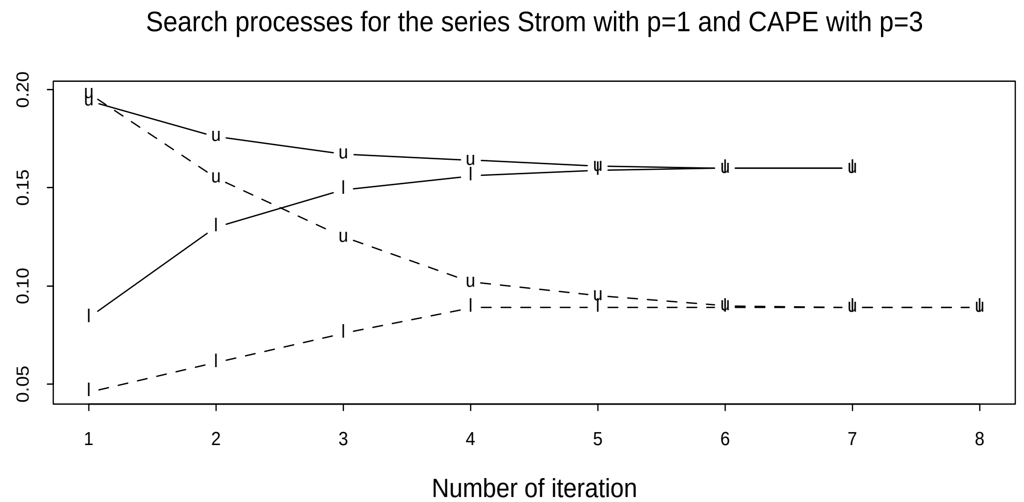


Figure 3: The search processes for the time series Strom with $p = 1$ (solid line) and CAPE with $p = 3$ (dashed line), where results with $h_0^1 = h_{\min}$ are marked by the letter "l" and those with $h_0^2 = h_{\max}$ are marked by the letter "u".

Härdle et al. (1988) proposed to estimate the MASE on $x \in [\Delta, 1 - \Delta]$ to avoid the boundary effect of a kernel estimator, where $\Delta > 0$ is a small positive number. In our proposal above no Δ (or $\Delta = 0$) is used, since local polynomial fitting has automatic boundary correction. However, if e.g. there are some outliers or there a structural change in the boundary area, then the estimate $\hat{g}^{(k)}$ depends strongly on the value of Δ. Hence we will use this idea as a diagnostic tool in order to see, if the selected bandwidth is susceptible to observations in the boundary area. The susceptibility of $\hat{h}$ to Δ is a signal, which shows that either the polynomial order p is not suitable or a global bandwidth does not work well for the given data set. As examples, bandwidths $\hat{h}^l$ selected for the two time series Strom and IFOR with h_0^1, $p = 1$ and 3 as well as $\Delta = 0.00, 0.02, \cdots, 0.10$ are given in Table 3. Note that if a $\Delta \neq 0$ is used, corresponding formulae given in Sections 2 to 4 should be adapted.

From Table 3 we see that the selected bandwidths in the two unusual cases, i.e. the series Strom with $p = 3$ and the series IFOR with $p = 1$ are very susceptible to boundary observations. The time series Strom seems to have some outliers at the right boundary. For different Δ, the influence of these outliers on $\hat{g}^{(4)}$ are quite different. This influence is however not so clear if $p = 1$ is used. Hence $p = 1$ should be chosen for smoothing this time series. For the time series IFOR, we see that there is a generally increasing trend until about the 145-th observation. The trend after there is more complex than before. This seems to be a structural change, which will be smoothed away, if $p = 1$ with a positive Δ is used. The selected bandwidth $\hat{h}^l$ with $\Delta = 0.10$ is about the same as $\hat{h}^r$ given in Table 2. Note that if Δ changes from 0.06 to 0.08, the selected bandwidth is

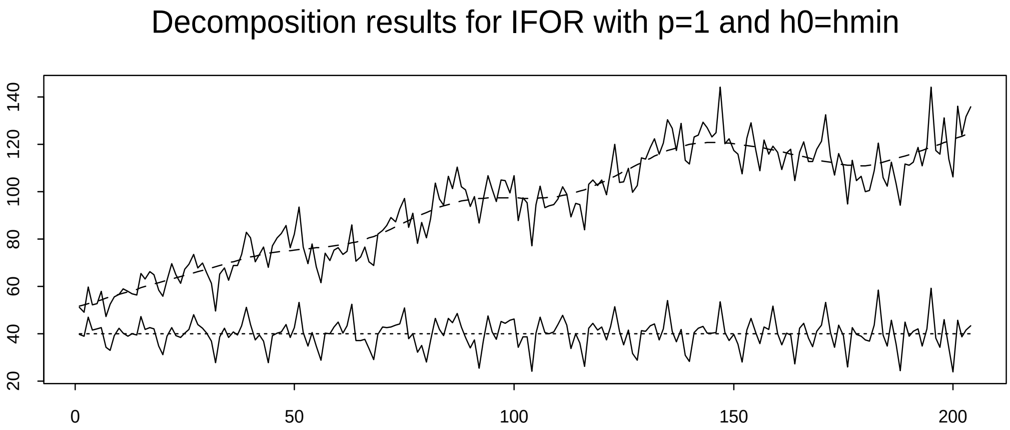
Table 3: $\hat{h}^l$ selected for Strom and IFOR with different Δ

Time Series	p	Δ = 0.00	0.02	0.04	0.06	0.08	0.10
Strom	1	0.160	0.161	0.163	0.163	0.163	0.163
	3	0.101	0.106	0.116	0.152	0.157	0.160
IFOR	1	0.113	0.118	0.128	0.141	0.259	0.256
	3	0.140	0.140	0.140	0.139	0.139	0.137

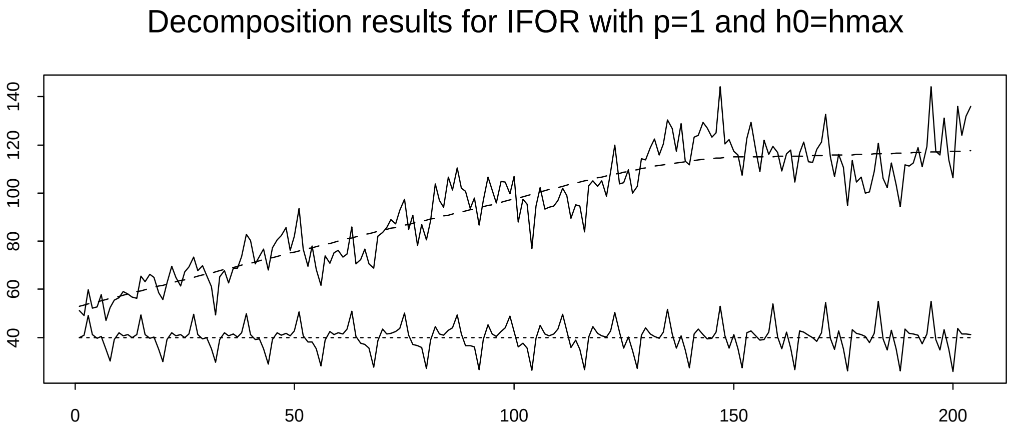
almost doubled. The structure of this time series can however be well fitted, if $p = 3$ is used. For the other two time series, especially for the series CAPE, both of $p = 1$ and $p = 3$ perform well.

In the following we will give an example for subjective choice from more than one fixed points obtained at the end of the procedure. Suppose that we would like to deal with the time series IFOR using $p = 1$. Then two stable fixed points, i.e. $h_f^1 = \hat{h}^l = 0.113$ and $h_f^2 = \hat{h}^r = 0.264$ as given in Table 2 will be found. Smoothing results for IFOR with $p = 1$ and these two bandwidths respectively are shown in Figure 4. We see that the results with $\hat{h}^r$ are clearly oversmoothed due to the reason mentioned above. Hence, for

$p = 1$, $\hat{h}^l$ should be chosen as the optimal bandwidth. The smoothing results with $\hat{h}^r$ also provide us some useful information about this data set.



(a)



(b)

Figure 4: Decomposition results for IFOR with $p = 1$ and different starting bandwidths. (a): results with h_0^1. (b): results with h_0^2. The curves are drawn similarly as in Figure 2.

6 Final remarks

In this paper an iterative plug-in algorithm for decomposing seasonal time series is developed. Simulated and data examples show that the proposed bandwidth selector performs well in practice. This proposal can also be applied to equidistant nonparametric regression without seasonality. In order to investigate the practical performance of the proposal in detail, a simulation study is required. This is however beyond the aim of this paper and will be carried out elsewhere.

An interesting question about bandwidth selection was raised by Prof. Gasser at a conference: "Should we choose a bandwidth subjectively or objectively?" Here we propose to carry out the procedure at least twice with $p = 1$ and $p = 3$. In cases when the results with $p = 1$ and $p = 3$ are both satisfactory, we can use the results with any p, e.g. those with $p = 3$. In this case the result may be considered to be *objective*. Sometimes we have however to choose p subjectively by means of our experience and more detailed analysis as shown in the last section. Furthermore, if there exist more than one stable fixed points for the p we would like to use, further subjective choice is also required. Note that both, the diagnosis at the boundary and the analysis of the smoothing results with a stable fixed point corresponding to a local minimum, will provide us more useful information about the structure of our data set. Hence, we recommend an experienced data analyst to use the proposal here in different ways.

The proposed bandwidth selector is motivated by optimizing $\hat{m} = \hat{g} + \hat{S}$. Similarly, we can develop an iterative plug-in algorithm for optimizing $\hat{g}$. Sometimes selection of optimal bandwidth for estimating S is also interesting. However, it is senseless to develop such a procedure under the assumption A3. If A3 holds, then a more preferable procedure is: 1. To estimate g with a corresponding optimal bandwidth and 2. To estimate S from the residuals $y_i - \hat{g}(x_i)$ parametrically, i.e. with the seasonal means.

In model (1) it is assumed that the errors are iid for simplicity. This is an impractical assumption, in particular for analyzing a time series. Recent works on iterative plug-in bandwidth selection in nonparametric regression with dependent errors (see Herrmann et al., 1992, Ray and Tsay, 1997 and Beran and Feng, 2002a, b) show that it is not difficulty to develop a data-driven procedure for model (1) with a general stationary time series error process.

Acknowledgements: This work was finished under the advice of Prof. Jan Beran, University of Konstanz, Germany, and was financially supported by the *Center of Finance and Econometrics* (CoFE) at the University of Konstanz. Some basic results used here are obtained in the author's PhD thesis, which was finished under the advice of Prof. Siegfried Heiler, University of Konstanz. The data for the time series CAPE and Hsales are downloaded from the *Time Series Data Library*. We would like to thank Prof. Rob J. Hyndman, Monash University, for making these data publicly available.

Appendix: Proofs of the results

A sketched proof of Lemma 1: The proof of this lemma based on some desirable standardizing and orthogonal finite sample properties of $\hat{g}$ and $\hat{S}$. These properties are quantified by the following properties of $\mathbf{w}_1$ and $\mathbf{w}_2$.

$$\begin{array}{lll}
a. & \sum_{i=1}^{n} w_{1i}(x_i - x_t)^j = \begin{cases} 1, & j = 0, \\ 0, & 1 \le j \le p, \end{cases} & \\
a'. & \begin{cases} \sum_{i=1}^{n} w_{1i}\cos(\lambda_j(i-t)) = 0, \\ \sum_{i=1}^{n} w_{1i}\sin(\lambda_j(i-t)) = 0, \end{cases} & j = 1, \ \dots \ , q. \\
b. & \sum_{i=1}^{n} w_{2i}(x_i - x_t)^j = 0, & 0 \le j \le p, \\
b'. & \begin{cases} \sum_{i=1}^{n} w_{2i}\cos(\lambda_j(i-t)) = 1, \\ \sum_{i=1}^{n} w_{i}\sin(\lambda_j(i-t)) = 0, \end{cases} & j = 1, \ \dots \ , q.
\end{array} \tag{A.1}$$

Note that $\mathbf{w}_1' = (\mathbf{e}_1', \mathbf{0}')(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}$ and $\mathbf{w}_2' = (\mathbf{0}', \Phi_\mathrm{s}')(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}$. Hence we have $\mathbf{w}_1'\mathbf{X} = (\mathbf{e}_1', \mathbf{0}')$ and $\mathbf{w}_2'\mathbf{X} = (\mathbf{0}', \Phi_\mathrm{s}')$. Observing the definition of $\mathbf{e}_1$ and Φ_s we obtain the results in (A.1). Note that (A.1) ensures that $\hat{g}$, $\hat{S}$ and hence $\hat{m}$ are exactly unbiased, if m is the sum of a polynomial of order no larger than p and S is an exactly periodic component with period s.

1. Under A2, A3 we have, in the neighbourhood of x_t,

$$g(x) = \sum_{j=0}^{p} \frac{g^{(j)}(x_t)}{j!}(x - x_t)^j + \frac{g^{(k)}(x_t + \theta(x - x_t))}{k!}(x - x_t)^k, \tag{A.2}$$

where $0 < \theta < 1$ and

$$S(x_i) = \sum_{j=1}^{q}(\beta_{2j}\cos\lambda_j(i-t) + [\beta_{3j}\sin\lambda_j(i-t)]). \tag{A.3}$$

This leads to $S(x_t) = \sum_{j=1}^{q} \beta_{2j}$. Following a', we have

$$B[\hat{g}(x_t)] = \sum_{i=1}^{n} w_{1i}[g(x_i) + S(x_i)] - g(x_t) = \sum_{i=1}^{n} w_{1i}g(x_i) - g(x_t), \tag{A.4}$$

since $\sum_{i=1}^{n} w_{1i}S(x_i) = 0$. For $B(\hat{S})$ we have

$$B[\hat{S}(x_t)] = \sum_{i=1}^{n} w_{2i}[g(x_i) + S(x_i)] - S(x_t) = \sum_{i=1}^{n} w_{2i}g(x_i), \tag{A.5}$$

since $\sum\limits_{i=1}^{n} w_{2i} S(x_i) = \sum\limits_{j=1}^{q} \beta_{2j} = S(x_t)$ following b' and (A.3). Property a' results in

$$\sum_{i=1}^{n} w_{2i} \left\{ \sum_{j=0}^{p} \frac{g^{(j)}(x_t)}{j!} (x - x_t)^j \right\} = 0. \tag{A.6}$$

Hence

$$\begin{aligned} B[\hat{S}(x_t)] &= \sum_{i=1}^{n} w_{2i} \frac{g^{(k)}(x_t + \theta(x_i - x_t))}{k!} (x_i - x_t)^k \\ &\doteq \frac{g^{(k)}(x_t)}{k!} h^k \sum_{i=1}^{n} w_{2i} \left(\frac{x_i - x_t}{h} \right)^k \\ &= o(h^k), \end{aligned} \tag{A.7}$$

where the last equation is due to the fact

$$\sum_{i=1}^{n} w_{2i} (\frac{x_i - x_t}{h})^{k'} = o(1), \text{ for any } k' \geq 0. \tag{A.8}$$

Equation (refss2uk) holds, since the weights w_{2i} are asymptotically periodic (see (7)). This shows that $B(\hat{S})$ is only due to the k-th order term in the Taylor expansion of g. And the contribution of this term to $B(\hat{S})$ is negligible compared with $B(\hat{g})$. We obtain

$$B[\hat{S}(x_t)] = o(B[\hat{g}(x_t)])$$

and

$$B[\hat{m}(x_t)] \doteq B[\hat{g}(x_t)].$$

Observe that $B(\hat{g})$ is the same as for a local polynomial fitting of order p, we obtain (9).

2. Detailed proof of (10) may be found in Feng (1999), where is it shown in particular that the two weighting systems $\mathbf{w}_1$ and $\mathbf{w}_2$ are asymptotically orthogonal in the sense that $\sum\limits_{i=1}^{n} w_{1i} w_{2i} = o(\sum\limits_{i=1}^{n} w_{1i}^2) = o(\sum\limits_{i=1}^{n} w_{2i}^2)$. This follows from (A.8), since $K_p(u)$ is a polynomial kernel.

3. Formula (11) follows from (9) and (10). Lemma 1 is proved. $\diamond$

In the following, it will be explained, why Lemma 2 and Theorem 1 should hold. Detailed proofs are omitted, since these results are similar to those in nonparametric regression without seasonality.

A sketched proof of Lemma 2: Case 1. Note that the two terms on the right hand side of (18) are due to the contribution of $B(\hat{g}^{(k)})$ and $\mathrm{var}(\hat{g}^{(k)})$ (see e.g. the proof of Proposition

1 in Beran and Feng, 2002a). In step 1 we have $h_{\mathrm{I},j} = o(h_{\mathrm{A}})$ in the j-th iteration. In this case $B(\hat{g}^{(k)})$ is negligible and $\hat{I}$ is dominated by $\mathrm{var}(\hat{g}^{(k)})$, which tends to infinite as $n \to \infty$. Observe that $w_i^k = O[(nh_{\mathrm{I},j}^{k+1})^{-1}]$, we have $\mathrm{var}(\hat{g}^{(k)}) = O(n^{-1}h_{\mathrm{I},j}^{-(2k+1)})$ and hence $\hat{I} = O_p(n^{-1}h_{\mathrm{I}j}^{-(2k+1)})$. Inserting this in the formula for h_j we obtain $h_j = O_p(h_{\mathrm{I}j})$, i.e. in this case h_{j-1} is inflated to a bandwidth of order $O_p(h_{\mathrm{I}j})$. Step 1 is proved. Results in Steps 2 and 3 are clear.

Case 2. Note that Step 1′ will not appear, if h_0 is of a larger order than h_{A} such that $h_0 \to 0$, since now A′1 is satisfied in the first iteration. In this case h_1 is already consistent and only Step 2′ will appear. Step 1′ occurs, only if $0 < h_0 < 0.5$ is taken to be a constant. Now $B(\hat{I}_1)$ is a constant and hence $\hat{I}_1 = O_p(1) = O_p(I)$. Now we obtain $h_1 = O_p(h_{\mathrm{A}})$, which is of the correct order but not yet consistent. The process will then be changed into Step 2′ in the second iteration. Lemma 2 is proved. $\diamond$

Remark A1. Theoretically, if the procedure is started with an h_0 such that $h_{\mathrm{A}} = o(h_0)$ and $h_0 \to 0$ as $n \to \infty$, then h_1 will already be consistent. Hence such a starting bandwidth is asymptotically more preferable. Now the asymptotic behaviour of an iterative plug-in bandwidth selector is easy to understand. If the sample size is small and the data have a special structure, a too large starting bandwidth, e.g. $h_0^2 = h_{\max}$ may perhaps lead to $\hat{I}_1 \doteq 0$. Now h_j could not be deflated to the optimal bandwidth. In the application we did not yet find such a phenomenon. If this occurs, it is no problem for our proposal, because it will be discovered by starting with the other bandwidth h_0^1.

A sketched proof of Theorem 1: The proof of Theorem 1 can be carried out based on a formula given in the appendix in Beran and Feng (2002a). See also Beran and Feng (2002b). They showed that, when convergence is reached, the rate of convergence of an iterative plug-in bandwidth selector is quantified by:

$$(\hat{h} - h_{\mathrm{A}})/h_{\mathrm{A}} \doteq -\frac{1}{2k+1-2\delta} I^{-1}(\hat{I} - I). \tag{A.9}$$

Equation (A.9) shows that $B(\hat{h})$ and $\mathrm{var}(\hat{h})$ at the end of the proposed procedure are of the corresponding orders as those of $\hat{I}$. $\mathrm{var}(\hat{h})$ is dominated by the second term in (19) of order $O(n^{-1}h_{\mathrm{I}}^{-4k-1})$, where h_{I} denotes the bandwidth for estimating I used at the end of the procedure, which is of order $O_p(n^{-1/7})$ for $k = 2$ and $O_p(n^{-1/13})$ for $k = 4$. In both cases, i.e. $k = 2$ with β_3 and $k = 4$ with β_2, the order of the second term on the right hand side of (18) is no larger than that of the first. Hence we have $B(\hat{h}) = O[B(\hat{I})] = O(h_{\mathrm{I}}^{-2})$. Straightforward calculation leads to the results of Theorem 1. $\diamond$

References

Beran, J. 1999. SEMIFAR models – A semiparametric framework for modelling trends, long range dependence and nonstationarity. Discussion paper, Center of Finance and Econometrics (CoFE), No. 99/16, Center of Finance and Econometrics, University of Konstanz.

Beran, J. and Feng, Y. (2002a). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *The Annals of the Institute of Statistical Mathematics* (in press).

Beran, J. and Feng, Y. (2002b). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. To appear in *Journal of Computational and Graphical Statistics.*

Beran, J., Feng, Y. and Heiler, S. (2000). Modifying the double smoothing bandwidth selection in nonparametric regression. Discussion Paper, CoFE, No. 00/37, University of Konstanz. Submitted.

Cleveland, W.S. 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. J. Amer. Statist. Assoc. 74, No. 36, 829–836.

Fan, J. and Gijbels, I. 1996. Local Polynomial Modeling and its Applications. Chapman & Hall, London.

Feng, Y. 1999. *Kernel- and Locally Weighted Regression* – with Application to Time Series Decomposition. Verlag für Wissenschaft und Forschung, Berlin.

Feng, Y. and Heiler, S. (2000). Eine robuste datengesteuerte Version des Berliner-Verfahrens (in German). *Wirtschaft und Statistik*, 10/2000, 786 – 795.

Gasser, T., Kneip, A. and Köhler, W. 1991. A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.*, 86, 643–652.

Gasser, T., Müller, H.G. and Mammitzsch, V. 1985. Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* 47 238–252.

Gasser, T., Sroka, L. and Jennen-Steinmetz, C. 1986. Residual Variance and Residual Pattern in Nonlinear Regression. Biometrika 73, 625–33.

Härdle, W., Hall, P. and Marron, J.S. 1988. How far are automatically chosen regression smoothing parameters from their optimum? (with discussion) *J. Amer. Statist. Assoc.*, 83, 86–99.

Hall, P., Kay, J.W. and Titterington, D.M. 1990. Asymptotically Optimal Difference-based Estimation of Variance in Nonparametric Regression. Biometrika 77, 521–8.

Heiler, S. 1966. Analyse der Struktur Wirtschaftlicher Prozesse durch Zerlegung von Zeitreihen. Dissertation, Tübingen.

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.

Heiler, S. and Feng, Y. 1996. Datengesteuerte Zerlegung saisonaler Zeitreihen. IFO-Studien 3/1996, 337–369.

Heiler, S. and Feng, Y. (2000). Data-driven decomposition of seasonal time series. *Journal of Statistical Planning and Inference*, 91, 351 – 363.

Heiler, S. and Michels, P. 1994. Deskriptive und Explorative Datenanalyse, Oldenbourg-Verlag, München.

Herrmann, E. and Gasser, T. 1994. Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. Preprint, Darmstadt Institute of Technology and University of Zürich.

Herrmann, E., Gasser, T. and Kneip, A. 1992. Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, 79, 783–795.

Makridakis, Wheelwright and Hyndman 1998. *Forecasting: Methods and Applications* (3rd edition). John Wiley, New York.

Müller, H.G. 1988. *Nonparametric Analysis of Longitudinal Data*, Springer-Verlag, Berlin.

Ray, B.K. and Tsay, R.S. 1997. Bandwidth selection for kernel regression with long-range dependence. *Biometrika*, 84, 791–802.

Rice, J. 1984. Bandwidth Choice for Nonparametric Regression. Annals of Statistics 12, 1215–30.

Ruppert, D., Sheather, S.J. and Wand, M.P. 1995. An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.* 90, 1257–1270.

Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.* **22** 1346–1370.

Schlittgen, R. and Streitberg, B. 1991. *Zeitreihenanalyse.* R. Oldenbourg, München.

Yuanhua Feng
Department of Mathematics and Statistics
University of Konstanz
D-78457 Konstanz, Germany

Email: yuanhua.feng@uni-konstanz.de
Tel. +49-7531-88-7363
Fax. +49-7531-88-2407