Center for International Economics

Working Paper Series

No. 2010-09

An iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method

Yuanhua Feng

November 2010

Center for International Economics
University of Paderborn
Warburger Strasse 100
33098 Paderborn / Germany

BUSINESS ADMINISTRATION & ECONOMICS
UNIVERSITY OF PADERBORN

An iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method

Yuanhua Feng

Faculty II - Business Administration and Economics

University of Paderborn, Warburger Str. 100, D-33098 Paderborn, Germany

Tel. 0049 5251 603379, Fax: 0049 5251 603955, E-Mail: yuanhua.feng@wiwi.upb.de

Abstract

We propose a fast data-driven procedure for decomposing seasonal time series using the Berlin Method, the software used by the German Federal Statistical Office in this context. Formula of the asymptotic optimal bandwidth h_{A} is obtained. Methods for estimating the unknowns in h_{A} are proposed. The algorithm is developed by adapting the well known iterative plug-in idea to time series decomposition. Asymptotic behaviour of the proposal is investigated. Some computational aspects are discussed in detail. Data example show that the proposal works very well in the practice and that data-driven bandwidth selection is a very useful tool to improve the Berlin Method. Deep insights into the iterative plug-in rule are also provided.

MSC2000 Codes: 62G08; 62G20; 62-07; 65Y15

Keywords: Time series decomposition, Berlin Method, local regression, bandwidth selection, iterative plug-in

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. In this paper the equidistant additive time series model

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

will be used to perform this, where $x_t = (t - 0.5)/n$, g is a smooth trend-cyclical component and S is a slowly changing seasonal component with period s. To simplify detailed discussion on bandwidth selection we assume in this paper that ϵ_t are iid random variables with $E(\epsilon_t) = 0$ and $\mathrm{var}(\epsilon_t) = \sigma^2$. The results can be easily extended to models with dependent errors. Model (1) can be treated as a nonparametric regression with an additional (deterministic) seasonal component. A traditional approach for estimating g and S is local regression with polynomials and trigonometric functions as local regressors (Heiler, 1970). This became the basis of the so-called Berlin Method (BV: Berliner Verfahren), which in its fourth version (BV4) is being used by the German Federal Statistical Office since 1983 (Speth, 2004, 2006). A great advantage of BV4 is its mathematical clarity. Hence BV4 is user-friendly (Cieplik, 2006). Moreover, it allows us to apply the recent developments in modern nonparametric regression to improve this approach.

A crucial problem by the use of BV4 is the selection of the bandwidth. This method performs better than related approaches only if the bandwidth is suitably selected. Heiler and Feng (2000) proposed to select the bandwidth under model (1) using a double-smoothing procedure. However, the running time of this procedure is very long, which hinders its application in the practice. In this paper a very fast and practically relevant algorithm for selecting bandwidth under model (1) is developed based on the iterative plug-in idea (Gasser et al., 1991). To our knowledge this is the first detailed study on plug-in bandwidth selection for time series decomposition. Moreover, results of this paper also provide deep insights into the iterative plug-in idea. Asymptotic behaviour of the proposal is investigated. Some computational aspects are discussed in detail. Application to different data example shows that the proposal works very well in the practice and that data-driven bandwidth selection is a very useful tool to improve the Berlin Method.

The paper is organized as follows. The estimators and related properties are described in Section 2. Estimation of the unknowns for bandwidth selection is discussed in Section 3. The plug-in algorithm is proposed and discussed in detail in Section 4. Data examples in Section 5 illustrate the practical usefulness of the proposal. Final remarks in Section 6 close the paper. 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. Denote by $m = g + S$ the mean function. A general version of BV4 is defined as follows. For more details see Feng (1999), and Heiler and Feng (2000). 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 (half) bandwidth. The locally weighted regression estimators of g, S and m at x_t are obtained by solving the least square problem

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

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 estimators 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. 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 observation vector and $\mathbf{K}$ denote a diagonal matrix with

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

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

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

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

and

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

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

The vectors $\mathbf{w} = (w_1, ..., w_n)'$, $\mathbf{w}_1 = (w_{11}, ..., w_{1n})'$ and $\mathbf{w}_2 = (w_{21}, ..., w_{2n})'$ are called weighting systems of $\hat{m}$, $\hat{g}$ and $\hat{S}$ respectively. 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

From here on it is assumed that p is odd so that $\hat{g}$ has automatic boundary correction. For the development of a plug-in bandwidth selector we need to discuss the asymptotic behaviour of $\hat{g}$, $\hat{S}$ and $\hat{m}$. 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 only made to avoid the estimation of the bias in $\hat{S}$. But model (1) works well in the case of slowly changing seasonality and a fixed selected bandwidth. Under A1 it can be showed that $\hat{g}$ is asymptotically equivalent to some kernel estimator. Hence the same asymptotic results in local polynomial fitting hold for $\hat{g}$ under model (1). The equivalent kernel for estimating g will be denoted by $K_p(u)$, 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 the mean averaged squared error (MASE) is used. 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. *The* MASE *of* $\hat{m}$ *is*

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

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 dominant part of the MASE is given by

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

where it is assumed that $I = \int \{g^{(k)}(x)\}^2 dx > 0$. The change in h_{A} due to S is just an additional term $(s-1) * R(K)$ in the kernel depending constant of the variance of $\hat{m}$. For $s = 1$ the above formulae reduce to known results in nonparametric regression (see e.g. 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 unknowns 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). This idea can be extended to seasonal-difference-based variance estimators under model (1) (see e.g Heiler and Feng, 2000). Here a sequence $D_{ms} = \{d_j|\ j = 0, 1, ..., m\}$ is called a *seasonal difference sequence*, if

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

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

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

Following Hall et al. (1990) it can be shown that under A2 and A3 $\hat{\sigma}_{\mathrm{D}}^2$ 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 we are only interested in the magnitude orders and these orders are the same for models with or without seasonality.

Assume now that

A1′. $h \to 0$ and $nh^{2k+1} \to \infty$ as $n \to \infty$.

A2′. g is at least l times continuously differentiable.

Under Assumptions A1′, A2′ 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

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

A1′ 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 $\hat{I}$ are at least consistent. See Ruppert et al. (1995) for related results in nonparametric regression without seasonality. The following remarks show 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 is smaller than that in Remark 1, but larger than that in Remark 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}. This order is smaller than both orders mentioned above.

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 an inflation method. Here an inflation method is a function $h_{\mathrm{I},j} = f(h_{j-1})$ such that $(h_{\mathrm{I},j})^{-1} = o[(h_{j-1})^{-1}]$. That is $h_{\mathrm{I},j}$ will be of a larger order than h_{A}, if h_{j-1} is at least of order $O(h_{\mathrm{A}})$. Now A1$'$ is satisfied so that $\hat{I}$ and $\hat{h}$ will be both consistent in the j-th iteration. Two inflation methods will be considered.

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

It is well known that the required number of iterations (J^0, say) by the MIM is very large, especially for $k > 2$. For example, if $k = 4$, it is $J^0 = 5k + 1 = 21$ for α_1 and $J^0 = (k + 1)(2k + 1) = 45$ for α_3 (see Herrmann and Gasser, 1994). Beran and Feng (2002a) introduced another inflation method $h_{\mathrm{I},j} = f(h_{j-1}) = ch^{\beta}_{j-1}$, called an exponential inflation method (EIM). This idea is studied by Beran and Feng (2002b) in detail. They 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. In the following the EIM with $c = 1$ will hence be used. Following Beran and Feng (2002b), the choices of β corresponding to α_1, α_2 and α_3 above 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 asymptotically optimal choice of β.

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 $j = 1, 2, ...$ 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

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

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

For related plug-in bandwidth selectors in nonparametric regression without seasonality see Gasser et al. (1991), Herrmann et al. (1992), Herrmann and Gasser (1994) and Rupper et al. (1995).

Theoretically, β_3 is the asymptotically optimal choice of β. Our experience show that, for $k = 2$, this choice works well in the 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 achieved. 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* A2′ *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 Herrmann and Gasser (1994, p. 8) and Beran and Feng (2002b). Note in particular that A1′ does not apply to Lemma 2.

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 reached 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 $\hat{h}$.

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 used. On the one hand, if $h_0 = o_p(1)$, then h_1 is already consistent, since A1′ 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 required 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.

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., 2009), which is of orders $O(n^{-2/5})$ for $k = 2$ and $O(n^{-2/9})$ for $k = 4$ and is hence negligible. Furthermore, the advantage of a plug-in bandwidth selector compared with a double-smoothing bandwidth selector is that it suns very fast. But the rate of convergence of a plug-in bandwidth selector is usually slower than a corresponding double-smoothing bandwidth selector (Feng and Heiler, 2009). This disadvantage is not so serious, because a slight change in the rate of convergence of a bandwidth selector will not affect the goodness-of-fit of the resulting nonparametric regression estimators.

4.3 Computational aspects

This subsection deals with such 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 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 determined 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 software 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 of 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 software the bandwidth actually used is an integer $b_h = [nh + 0.5]$, which is the (half) bandwidth w.r.t. the observation time t. The total number of observations used at each time point is $N_h = 2b_h + 1$. Let $b_{h_{\mathrm{I},j}} = [nh_{\mathrm{I},j} + 0.5]$. Then we obtain a natural criterion for stopping the computing procedure, i.e. the procedure will be stopped, if $b_{h_{\mathrm{I},j^0}} = b_{h_{\mathrm{I},j^0-1}}$ 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 minima 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 by $\hat{h}^r$ 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 will 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 fixed points. In the case when the finite MASE has more than one local minima, 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 minima. If this is the case, we should find out all possible stable fixed points and then select one of them as the bandwidth to use by analyzing the smoothing results further.

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

The following data examples are chosen to show the practical performance of the proposal.

1. The 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. The 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. The 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. The Series "Hsales" – Monthly sales of new one-family houses sold in the USA from 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. Throughout the application the bisquare kernel is used. The selected bandwidths and the number of iterations with the smallest starting bandwidth $h_0^1 = h_{\min} = s/n$ and the largest starting bandwidth $h_0^2 = h_{\max} = 0.5 - 1/n$, together with the answer, if the two bandwidth are the same or not, are listed in Table 1 for all data examples. From Table 1 we see that the two selected bandwidths 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 should be mentioned: 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 $p = 1$. This means that the proposal does not work well for Strom series with $p = 3$. For the final smoothing we hence propose to use $p = 1$ for Strom series and

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

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

$p = 3$ for the others. Data-driven decomposition results for these examples are shown in Figures 1a through 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 1 look quite well. This shows the practical usefulness of the proposed procedure.

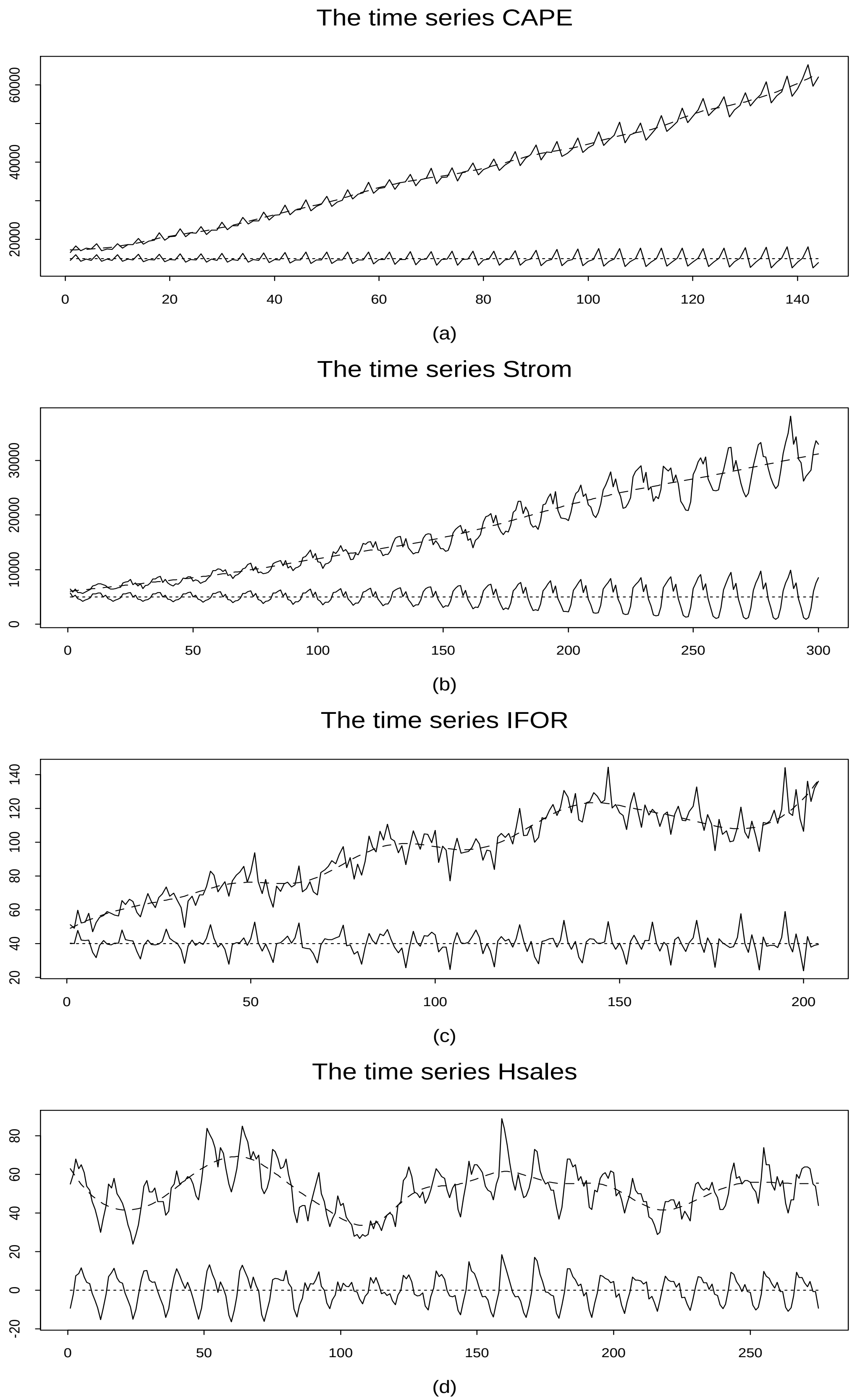

Figure 1: Optimal decomposition results for the data examples. Upper: the data together with the estimated trend (dashes). Below: the estimated seasonal component.

Note that the selected bandwidths for the examples given in Figure 1a through 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 $s = 12$. Table 1 also shows that j^0 changes from case to case.

Furthermore, it is easy to calculate that the above bandwidths for the four examples correspond to 27 seasons, and 97, 57, 59 months, respectively. The selected bandwidths in the first two cases are much larger than the bandwidths used in the current version of BV4 (BV4.1). See Speth (2004, 2006). This shows that the performance of BV4.1 may be improved clearly, if it can be combined with the proposed bandwidth selection algorithm.

Finally, we want to show some detailed properties of the iterative plug-in algorithm so that the reader can understand the proposal well. Following Lemma 2 we have $\hat{h}^l \leq h_{\mathrm{A}} \leq \hat{h}^r$ in probability. From Table 1 we see that this is true for all 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 2, where the results are for the time series Strom with $p = 1$ (solid line) and CAPE with $p = 3$ (dashed line). From Figure 2 we can see although the selected bandwidth for Strom with $p = 1$ is much larger than

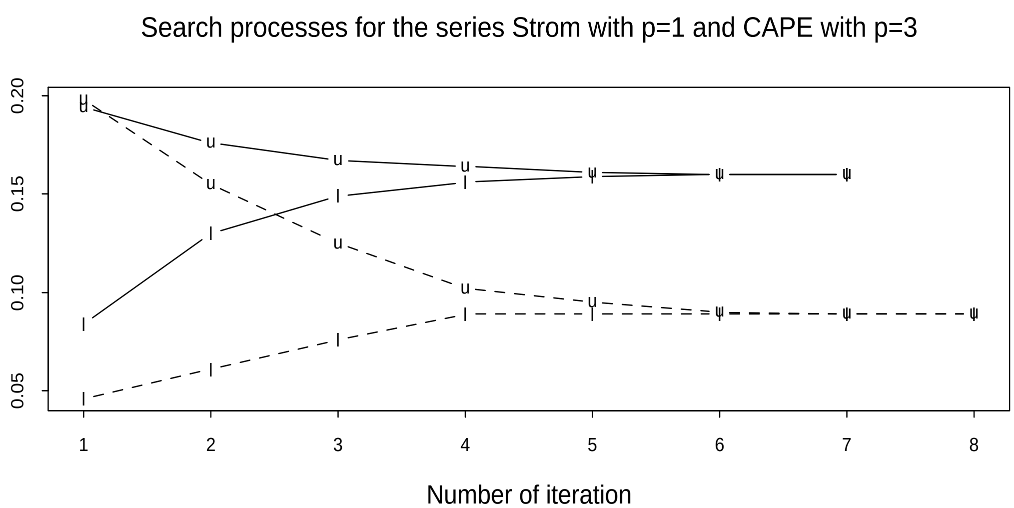

Figure 2: Search processes for Strom ($p = 1$, solid line) and CAPE ($p = 3$, dashed line). The letters "l" and "u" indicate results with $h_0^1 = h_{\min}$ and $h_0^2 = h_{\max}$, respectively.

that for CAPE with $p = 3$, h_1 with h_0^2 in the second case is however slightly larger than that in the first case. But, after a few iterations both of them achieve their corresponding fixed points.

6 Final remarks

This paper proposes an iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method. Computational aspects of the proposal are discussed in detail. A few nice properties of the iterative-plug rule were found. Data examples show that the proposal works very well in the practice. The facts that the selected bandwidths vary from one series to another very strongly and that all of the selected bandwidths are clearly larger than the default bandwidth used in BV4 indicate that the introduction of a suitable bandwidth selector into the current BV4 is necessary. This will help to improve the quality of this software clearly. This study is the first detailed study on bandwidth selection for decomposing economic time series. There are still quite a lot open questions in this context. For instance, the proper combination of the proposed algorithm with BV4.1, the adaptation of the algorithm according to the dependence structure of the errors, bandwidth selection under consideration of the bias in $\hat{S}$ and selection of two separate bandwidths for estimating g and S, respectively.

Acknowledgements: This work was partly supported by the *Center for International Economics*, University of Paderborn. 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 $\mathbf{w}_2$.

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

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 and 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 $\sum_{i=1}^{n} w_{2i}S(x_i) = \sum_{j=1}^{q} \beta_{2j} = S(x_t)$ 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 (A.8) 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 $\sum_{i=1}^{n} w_{1i}w_{2i} = o(\sum_{i=1}^{n} w_{1i}^2) = o(\sum_{i=1}^{n} w_{2i}^2)$. 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 A1′ 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. ◇

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. ◇

References

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*, 54, 291-311.

Beran, J. and Feng, Y. (2002b). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. *J. Computat. Graph. Statist.*, 11, 690-713.

Beran, J., Feng, Y. and Heiler, S. (2009). Modifying the double smoothing bandwidth selector in nonparametric regression. *Statist. Methodology*, 6, 447-65.

Cieplik, U. (2006). BV4.1 Methodology and User-friendly Software for Decomposing Economic Time Series. German Federal Statistical Office.

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. (2009). A simple bootstrap bandwidth selector for local polynomial fitting. *J. Statist. Comput. and Simul.*, 79, 1425-39.

Gasser, T., Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.*, 86, 643-52.

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

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. (1970). Theoretische Grundlagen des "Berliner Verfahrens". In Wetzel, W. (Ed.): Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse. *Allg. Statistischen Archiv*, Sonderheft 1, 67-93.

Heiler, S. and Feng, Y. (2000). Data-driven decomposition of seasonal time series. *J. Statist. Pl. Inf.*, 91, 351-63.

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, 783-95.

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

Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *Ann. Statist.*, 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, 1257-70.

Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *Ann. Statist.*, 22 1346-70.

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

Speth, Hans-Theo (2004). Komponentenzerlegung und Saisonbereinigung ökonomischer Zeitreihen mit dem Verfahren BV4.1. German Federal Statistical Office, Methodenberichte 3.

Speth, H.-Th. (2006). The BV4.1 Procedure for Decomposing and Seasonally Adjusting Economic Time Series (English translation of Speth, 2004). German Federal Statistical Office.

Recent discussion papers

2010-09	Yuanhua Feng	An iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method
2010-08	Zhichao Guo Yuanhua Feng Xiangyong Tan	Short- and long-term impact of remarkable economic events on the growth causes of China-Germany trade in agri-food products
2010-07	B. Michael Gilroy Elmar Lukas Christian Heimann	Welchen Einfluss hat die Anwesenheit von ausländischen und multinationalen Unternehmungen auf die deutschen Exporte?
2010-06	Stefan Gravemeyer Thomas Gries	Income and disparity in Germany and China
2010-05	Thomas Gries Margarete Redlin	Short-run and Long-run Dynamics of Growth, Inequality and Poverty in the Developing World
2010-04	Stefan Gravemeyer Thomas Gries Jinjun Xue	Poverty in Shenzen
2010-03	Alexander Haupt Tim Krieger Thomas Lange	A Note on Brain Gain and Brain Drain: Permanent Migration and Education Policy
2010-02	Sarah Brockhoff TimKrieger Daniel Meierrieks	Ties That Do Not Bind (Directly): The Education Terrorism Nexus Revisited
2010-01	Claus-Jochen Haake, Tim Krieger, Steffen Minter	On the institutional design of burden sharing when financing external border enforcement in the EU
2009-06	Tim Krieger, Stefan Traub	Wie hat sich die intragenerationale Umverteilung in der staatlichen Säule des Rentensystems verändert? Ein internationaler Vergleich auf Basis von LIS-Daten
2009-05	Karin Mayr, Steffen Minter, Tim Krieger	Policies on illegal immigration in a federation
2009-04	Tim Krieger, Daniel Meierrieks	Terrorism in the Worlds of Welfare Capitalism *[forthcoming in: Journal of Conflict Resolution]*
2009-03	Alexander Haupt, Tim Krieger	The role of mobility in tax and subsidy competition
2009-02	Thomas Gries, Tim Krieger, Daniel Meierrieks	Causal Linkages Between Domestic Terrorism and Economic Growth *[forthcoming in: Defense and Peace Economics]*
2009-01	Andreas Freytag, Jens J. Krüger, Daniel Meierrieks, Friedrich Schneider	The Origin of Terrorism - Cross-Country Estimates on Socio-Economic Determinants of Terrorism
2008-11	Thomas Gries, Magarete Redlin	China's provincial disparities and the determinants of provincial inequality

		[published in: Journal of Chinese economic and business studies 7 (2009), 2, 259-281]
2008-10	Thomas Gries, Manfred Kraft, Daniel Meierrieks	Financial Deepening, Trade Openness and Economic Growth in Latin America and the Caribbean *[forthcoming in: Applied Economics]*
2008-09	Stefan Gravemeyer, Thomas Gries, Jinjun Xue	Discrimination, Income Determination and Inequality – The case of Shenzen *[forthcoming in: Urban Studies]*
2008-08	Thomas Gries, Manfred Kraft, Daniel Meierrieks	Linkages between Financial Deepening, Trade Openess and Economic Development: Causality Evidence from Sub-Saharan Africa *[published in: World Development 37 (2009), 1849-1860]*
2008-07	Tim Krieger, Sven Stöwhase	Diskretionäre rentenpolitische Maßnahmen und die Entwicklung des Rentenwerts in Deutschland von 2003-2008 *[published in: Zeitschrift für Wirtschaftspolitik 58 (2009), 1, 36-54]*
2008-06	Tim Krieger, Stefan Traub	Back to Bismarck? Shifting Preferences for Intragenerational Redistribution in OECD Pension Systems
2008-05	Tim Krieger, Daniel Meierrieks	What causes terrorism? *[forthcoming in: Public Choice]*
2008-04	Thomas Lange	Local public funding of higher education when students and skilled workers are mobile *[published in: Finanzarchiv 65 (2009), 2, 178-199]*
2008-03	Natasha Bilkic, Thomas Gries, Margarethe Pilichowski	Stay at school or start working? - Optimal timing of leaving school under uncertainty and irreversibility
2008-02	Thomas Gries, Stefan Jungblut, Tim Krieger, Henning Meier	Statutory retirement age and lifelong learning
2008-01	Tim Krieger, Thomas Lange	Education policy and tax competition with imperfect student and labor mobility [forthcoming in: International Tax and Public Finance]
2007-05	Wolfgang Eggert, Tim Krieger, Volker Meier	Education, unemployment and migration *[published in: Journal of Public Economics 94 (2010), 5-6, 354-362.]*
2007-04	Tim Krieger, Steffen Minter	Immigration amnesties in the southern EU member states - a challenge for the entire EU? *[published in: Romanian Journal of European Studies 5-6/2007, 15-32.]*
2007-03	Axel Dreher, Tim Krieger	Diesel price convergence and mineral oil taxation in Europe *[published in: Applied Economics 42 (2010), 15, 1955-1961]*
2007-02	Michael Gorski, Tim Krieger, Thomas Lange	Pensions, education and life expectancy

2007-01	Wolfgang Eggert, Max von Ehrlich, Robert Fenge, Günther König	Konvergenz- und Wachstumseffekte der europäischen Regionalpolitik in Deutschland *[published in: Perspektiven der Wirtschaftspolitik 8 (2007), 130-146.]*
2006-02	Tim Krieger	Public pensions and return migration *[published in: Public Choice 134 (2008), 3-4, 163-178.]*
2006-01	Jeremy S.S. Edwards, Wolfgang Eggert, Alfons J. Weichenrieder	The measurement of firm ownership and its effect on managerial pay *[published under the title "Corporate Governance and Pay for Performance: Evidence from Germany" in: Economics of Governance 10 (2009), 1, 1-26.]*