Modifying the double smoothing bandwidth selector in nonparametric regression*

by

Jan Beran, Yuanhua Feng and Siegfried Heiler

University of Konstanz

Abstract

In this paper a modified double smoothing bandwidth selector, $\hat{h}_{\mathrm{MDS}}$, based on a new criterion, which combines the plug-in and the double smoothing ideas, is proposed. A self-complete iterative double smoothing rule ($\hat{h}_{\mathrm{IDS}}$) is introduced as a pilot method. The asymptotic properties of both $\hat{h}_{\mathrm{IDS}}$ and $\hat{h}_{\mathrm{MDS}}$ are investigated. It is shown that $\hat{h}_{\mathrm{MDS}}$ performs asymptotically very well. Moreover, it is asymptotically negatively correlated with h_{ASE}, the minimizer of the averaged squared error. The asymptotic performances of $\hat{h}_{\mathrm{MDS}}$ and of the iterative plug-in method, $\hat{h}_{\mathrm{IPL}}$ (Gasser et al., 1991) are compared. A comparative simulation study is carried out to show the practical performance of $\hat{h}_{\mathrm{MDS}}$ and related methods. It is shown that $\hat{h}_{\mathrm{MDS}}$ seems to be the best in the practice. Finite sample negative correlations between the chosen bandwidth selectors and h_{ASE} are also studied.

Key Words: Bandwidth selection, double smoothing, nonparametric regression, plug-in.

*This work was supported by the Center of Finance and Econometrics (CoFE) at the University of Konstanz.

1 Introduction

Consider the equidistant fixed design nonparametric regression model

$$Y_i = m(x_i) + \epsilon_i, \tag{1.1}$$

where $x_i = (i - 0.5)/n$, h is the bandwidth and ϵ_i are iid errors with mean zero and variance σ^2. Our goal is to estimate the curve $m(\cdot)$ from these n observations. In this paper we use the Nadaraja-Watson kernel estimator

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K[(x - x_i)/h]Y_i}{\sum_{i=1}^{n} K[(x - x_i)/h]} =: \sum_{i=1}^{n} w_{ih}(x)Y_i, \tag{1.2}$$

where K is a kernel of order r (see Gasser, Müller and Mammitzsch, 1985) and h is the bandwidth. For non-equidistant designs the Gasser-Müller estimator (Gasser and Müller, 1994) is preferable.

The practical performance of $\hat{m}_h$ depends strongly on the bandwidth h. Various procedures of bandwidth selection have been proposed in the statistical literature. All of the classical methods (see Härdle, Hall and Marron, 1988 for a survey) are known to be subjected to an unacceptably large amount of sample variation. In recent years, some modern bandwidth selectors, which perform well both theoretically and in practice, were proposed. Two important ideas are the plug-in (PL) rule (Gasser, Kneip and Köhler, 1991 and Ruppert, Sheather and Want, 1995) and the double smoothing (DS) procedure (Müller, 1985, Härdle, Hall and Marron, 1992, Heiler and Feng, 1998, Feng, 1999 and Feng and Heiler, 1999). Other proposals may be found e.g. in Chiu (1991) and Fan and Gibels (1995). This paper focuses on improving the existing methods for selecting a global bandwidth h.

1.1 Criteria of assessing the performance

There are two widely used measures for assessing the performance of $\hat{m}_h$, namely the averaged squared error (ASE)

$$\Delta(h) = n^{-1} \sum_{i}^{*} [\hat{m}_h(x_i) - m(x)]^2 \tag{1.3}$$

and its mean, the mean averaged squared error (MASE)

$$M(h) = E[\Delta(h)] = n^{-1} \sum_{i}^{*} E[\hat{m}_h(x_i) - m(x)]^2, \tag{1.4}$$

where $\sum_i^*$ denotes summation over indices i such that $c < x_i < d$, where $0 < c < d < 1$ are introduced to reduce the boundary effects of a kernel estimate. Denoted by h_{ASE} and h_{M} the minimizers of ASE and MASE, respectively. Both, h_{ASE} and h_{M}, can be considered as "optimal bandwidth" in some sense. Note that h_{ASE} is itself a random variable. To design a bandwidth selector that is less sensitive to the sample variation, h_{M} rather than h_{ASE} should be used as the target. The reason is that h_{M} can be estimated with the highest relative rate of convergence $n^{-1/2}$ under standard conditions. However, h_{ASE} cannot be estimated with a relative rate of convergence higher than $n^{-1/(2(2r+1))}$, which is $n^{-1/10}$ for $r = 2$, no matter how many derivatives are assumed to exist. Even the difference between h_{M} and h_{ASE} is of size $n^{-3/(2(2r+1))}$ (i.e. of the relative order $n^{-1/(2(2r+1))}$). In fact we have

$$n^{3/(2(2r+1))}(h_{\mathrm{ASE}} - h_{\mathrm{M}}) \longrightarrow N(0, \sigma_1^2) \tag{1.5}$$

in distribution (see Härdle et al., 1988), where σ_1^2 is the same as as the σ_2^2 defined in Härdle et al. (1988).

In principle, h_{ASE} (and not h_{M}) should be called the "optimal bandwidth", since it makes $\hat{m}_h$ as close as possible to m for the data set at hand, instead of for the average over all possible data sets. Observing however that, h_{M} also performs quite well (although it is not efficient following Hall and Johnstone, 1992), each of the modern bandwidth selectors attempts to come close to the good performance of h_{M} instead of estimating h_{ASE}. Fortunately, many simulation results show that all of the recently proposed bandwidth selectors perform clearly better than the classical ones, not only in terms of h_{M} but also in terms of h_{ASE}. In this paper, h_{M} will be taken as the target and will be called the optimal bandwidth. However, the practical performance of a bandwidth selector will be assessed following the ASE, or equivalently following its distance to h_{ASE}.

1.2 Motivation

Obviously, in the commonly used case with $r = 2$, any bandwidth selector $\hat{h}$ that comes within $o_p(n^{-3/10})$ of h_{M} will have the asymptotic property as given in (1.5), i.e.

$$n^{3/10}(h_{\mathrm{ASE}} - \hat{h}) \longrightarrow N(0, \sigma_1^2) \tag{1.6}$$

in distribution. Observing that the difference between almost all of the recently proposed bandwidth selectors and h_{M} is of size $o_p(n^{-3/10})$, it is worthless to assess them w.r. to h_{ASE} asymptotically, since they are now all asymptotically equivalent. However, these bandwidth selectors may perform quite differently when compared asymptotically with h_{M}. The goal of this paper is to propose a modified DS bandwidth selector, which has good asymptotic properties w.r. to h_{M} and at the same time performs well for finite sample in term of h_{ASE}.

There are some reasons for choosing the DS rather than the plug-in rule: 1. This method does not require the use of the asymptotic formula for the bias part in MASE and hence does not involve the estimation of m'' (in case of $r = 2$); 2. The DS idea is a very flexible bandwidth selection rule. There are many variants of it (see Härdle et al. 1992 and Heiler and Feng, 1998). And it can also be easily used for selecting bandwidth for estimation of derivatives (see Müller, 1985); 3. Asymptotic properties of it are often superior to those of a PL method under given conditions; 4. If the bandwidth is selected on the whole support $[0, 1]$, then the so-called boundary effect will play a more serious role for a plug-in bandwidth selector than for a DS bandwidth selector. Furthermore, in many cases when a plug-in method is not well defined and hence is not asymptotically optimal (see e.g. Gasser et al., 1991 for some examples), the DS bandwidth selector may still be optimal. Here a bandwidth selector is said to be asymptotically optimal, if $\hat{h}/h_{\mathrm{M}} \to 1$ in probability as $n \to \infty$.

The DS bandwidth selectors proposed so far use the exact formula for estimating the variance. This makes the method unnecessarily complex. Another problem is that, like for the PL method, we need a method to select the bandwidth at the pilot stage. In the proposal by Feng and Heiler (1999), denoted by $\hat{h}_{\mathrm{ODS}}$, the R-criterion (Rice, 1984) is used as the pilot method. Such a simple DS rule works well but is not yet satisfactory (see the simulation in section 4). In this paper, the bandwidth selector $\hat{h}_{\mathrm{ODS}}$ is modified in two ways. At first, the estimate is simplified by introducing the use of the asymptotic formula for the variance part of MASE (like for a PL method). Then an iterative double smoothing (IDS) bandwidth selector, $\hat{h}_{\mathrm{IDS}}$, which is related to the iterative plug-in (IPL) method, $\hat{h}_{\mathrm{IPL}}$ (Gasser et al., 1991), is proposed and used as the pilot method. This makes the DS method self-complete. The finite sample performance is also improved by using the DS based pilot procedure. These two improvements together make it possible to extend the DS idea to nonparametric regression with short- or long-range dependent data. This is indeed the original motivation of this study, which will however not be discussed in this paper.

1.3 Summary and organization

The modified double smoothing (MDS) criterion, the pilot method $\hat{h}_{\mathrm{IDS}}$ and the main proposal, $\hat{h}_{\mathrm{MDS}}$, are defined in section 2 after a brief description of the DS and the PL ideas. The asymptotic properties of $\hat{h}_{\mathrm{IDS}}$ and of $\hat{h}_{\mathrm{MDS}}$ are investigated in section 3. It is shown that, although $\hat{h}_{\mathrm{MDS}}$ performs asymptotically very well, it is still asymptotically negatively correlated with h_{ASE}. Note that the latter is also the case for $\hat{h}_{\mathrm{IPL}}$ (see Herrmann, 1994). The results in Theorem 2 allow us to compare the asymptotic performances of $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{IPL}}$. It is also explained why $\hat{h}_{\mathrm{MDS}}$ would perform better in practice than $\hat{h}_{\mathrm{ODS}}$. Section 4 summarizes the results of a comparative simulation study. It is shown that, $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{ODS}}$ perform quite differently in practice, especially when n is small, although their asymptotic properties are almost the same. It is also shown that $\hat{h}_{\mathrm{MDS}}$ performs clearly better than $\hat{h}_{\mathrm{IPL}}$ in some cases. Furthermore, it is shown that all of the selected bandwidth selectors are clearly negatively correlated with h_{ASE}. The simulation study confirms the theoretical results. Some concluding remarks are put in section 5. Proofs of results are given in the appendix.

2 The proposals

In the following basic ideas for the proposals in this paper will be described.

2.1 The double smoothing idea

The DS idea was first introduced in the statistical literature by Gasser, Müller, Köhler, Molinari and Prader (1984) and its properties were then discussed in Müller (1985). This approach focuses on minimizing a direct estimate of $M(h)$. Note that $M(h)$ splits into a variance and a bias part, i.e. $M(h) = V(h) + B(h)$ with

$$\begin{aligned} V(h) &= n^{-1} \sum_i^{*} \mathrm{var}\left[\hat{m}(x_i)\right] \\ &= n^{-1}\sigma^2 \sum_i^{*} \sum_{j=1}^{n} w_{jh}(x_i)^2 \end{aligned}$$

and

$$B(h) = n^{-1} \sum_i^{*} b(x_i)^2$$

$$= \quad n^{-1} \sum_i^* \{E[\hat{m}(x_i)] - m(x_i)\}^2.$$

Let $\hat{\sigma}^2$ be the variance estimator proposed by Gasser, Sroka, and Jennen-Steinmetz (1986) defined by

$$\hat{\sigma}^2 = \frac{2}{3(n-2)} \sum_{i=1}^{n-2} (Y_i - \frac{1}{2} Y_{i-1} - \frac{1}{2} Y_{i+1})^2. \tag{2.1}$$

The variance part of $M(h)$ can be estimated by

$$\hat{V}(h) = n^{-1} \hat{\sigma}^2 \sum_i^* \sum_{j=1}^n w_{jh}(x_i)^2. \tag{2.2}$$

Following the DS idea, the bias is estimated by means of a pilot estimate with a kernel L of order s and another bandwidth g:

$$\hat{m}_g(x) = \frac{\sum_{i=1}^n L[(x-x_i)/g] Y_i}{\sum_{i=1}^n L[(x-x_i)/g]} =: \sum_{i=1}^n w_{ig}(x) Y_i, \tag{2.3}$$

where w_{ig} $(i = 1, 2, ..., n)$ denote the weights for the pilot estimate and L is allowed to be different from K. The bias part of M is now estimated by

$$\begin{aligned} \hat{B}(h,g) &= n^{-1} \sum_i^* \hat{b}(x_i)^2 \\ &= n^{-1} \sum_i^* \left\{ \sum_{j=1}^n w_{hj}(x_i) \hat{m}_g(x_j) - \hat{m}_g(x_i) \right\}^2 . \end{aligned} \tag{2.4}$$

Note that, $\hat{B}(h,g)$ is obtained from $B(h)$ by replacing $m(h)$ by its estimate. Following Feng and Heiler (1999), $\hat{B}(h,g)$ may be interpreted as a bootstrap bias estimator. The final (ordinary) DS estimator of $M(h)$ is defined by

$$\hat{M}(h,g) = \hat{V}(h) + \hat{B}(h,g). \tag{2.5}$$

An ordinary DS bandwidth selector is defined as the minimizer of $\hat{M}(h,g)$. Definition (2.5) directly follows the proposal in Müller (1985). Härdle et al. (1992) introduced a slightly different DS criterion. Heiler and Feng (1998) proposed to combine these two definitions in a unified approach and introduced the use of a factorized pilot bandwidth. In this paper only a fixed pilot bandwidth will be considered. In order that a DS procedure is data-driven, we need to have a proper data-driven procedure for selecting g. This will be investigated in subsection 2.4. In this paper it is assumed that r and s are both even and $s \geq r$.

2.2 The plug-in method

For a kernel function of order r define

$$R(K) = \int_{-1}^{1} K(u)^2 du \qquad \text{and} \qquad \kappa_r = \frac{1}{r!}\int_{-1}^{1} u^r K(u) du,$$

where κ_r is called the kernel constant of K. And for the regression function m, which is assumed to be at least r time continuously differentiable, define

$$I(m^{(r)}) = \int_c^d \{m(x)^{(r)}\}^2 dx.$$

Then we have an approximation of $M(h)$

$$\begin{aligned} M_{\mathrm{A}}(h) &= V_{\mathrm{A}}(h) + B_{\mathrm{A}}(h) \\ &= \frac{\sigma^2}{nh} R(K)(d-c) + h^{2r}\kappa_r^2 I(m^{(r)}). \end{aligned} \tag{2.6}$$

The asymptotically optimal bandwidth, which minimizes the AMASE, is

$$h_{\mathrm{A}} = c_0 n^{-1/(2r+1)} \tag{2.7}$$

with

$$c_0 = \left(\frac{(d-c)\sigma^2 R(K)}{2r\,\kappa_r^2\, I(m^{(r)})} \right)^{1/(2r+1)}. \tag{2.8}$$

A PL bandwidth selector is obtained by replacing the unknowns, σ^2 and $I(m^{(r)})$, in (2.8) by consistent estimates.

It is well known that $\hat{\sigma}^2$ defined in (2.1) is a root n consistent estimator of σ^2. Hence the key problem here is to estimate $I(m^{(r)})$. A natural estimate of $I(m^{(r)})$ is

$$\hat{I}(m^{(r)}) = n^{-1} \sum_i^{*} \{\hat{m}^{(r)}(x_i; b)\}^2, \tag{2.9}$$

where $\hat{m}^{(r)}(x_i; b)$ is a kernel estimate of $m^{(r)}$ based on a kernel for estimating the r-th derivative (see Gasser et al., 1985) and a bandwidth b. Again, we need to select the pilot bandwidth b for estimating $m^{(r)}$.

The IPL procedure (for $r = 2$) proposed by Gasser et al. (1991) is motivated by fixpoint search. Their proposal, $\hat{h}_{\mathrm{IPL}}$, proceeds as follows

1. Begin with the smallest bandwidth $h_0 = 1/n$;

2. Estimate $\hat{h}_i(m'')$ in the i-th iteration with the bandwidth $b_i = h_{i-1} * n^{1/10}$;

3. Calculate $\hat{h}_i$ following (2.7);

4. Stop at the 11-th iteration and put $\hat{h}_{\mathrm{IPL}} = \hat{h}_{11}$.

For details see Gasser et al. (1991) and Herrmann (1994). Some improvements of the original proposal in Gasser et al. (1991) may be found in Herrmann and Gasser (1994). Beran (1999) proposed the use of an exponential inflation method in the IPL procedure, which is discussed in detail in Beran and Feng (1999, 2000). Advantages of the IPL idea are its stability and simple generalization to bandwidth selection in nonparametric regression with dependent data. See Herrmann, Gasser and Kneip (1992) for an IPL bandwidth selector for data with short-range dependence and Ray and Tsay (1997) and Beran and Feng(1999, 2000) for data with long-range dependence. Another data-driven PL procedure may be found in Ruppert et al. (1995).

2.3 The MDS criterion

Note that, the key point of the DS rule is the bootstrap estimate of $B(h)$ not the estimate of $V(h)$. $\hat{V}(h)$ in (2.2) does not involve the pilot estimate $\hat{m}_g$. It does even not depend on the unknown function m anymore. However, (2.2) depends strongly on the iid assumption. It is rather difficult to extend (2.2) and hence this idea to the context of nonparametric regression with dependent errors. Hence we propose to estimate the variance part of $M(h)$ using the much simpler asymptotic formula $V_{\mathrm{A}}(h)$ rather than $V(h)$. By doing this we obtain a MDS estimator of $M(h)$

$$\hat{M}_{\mathrm{M}}(h, g) = \hat{V}_{\mathrm{A}}(h) + \hat{B}(h, g). \tag{2.10}$$

Now, a MDS bandwidth selector is defined as the minimizer of $\hat{M}_{\mathrm{M}}$ in (2.10). Although $\hat{M}_{\mathrm{M}}$ is obtained by combining the PL and the DS ideas, a MDS bandwidth selector does not share the disadvantages of the PL method.

Indeed, the use of $\hat{M}_{\mathrm{M}}$ instead of $\hat{M}$ does not cause any clear loss in accuracy of the selected bandwidth. The basis for this conclusion is that, asymptotically, the difference between $M(h_{\mathrm{M}})$ and $M_{\mathrm{A}}(h_{\mathrm{M}})$ is dominated by the approximation in $B(h_{\mathrm{M}})$, i.e. $M(h_{\mathrm{M}}) - M_{\mathrm{A}}(h_{\mathrm{M}}) \doteq B(h_{\mathrm{M}}) - B_{\mathrm{A}}(h_{\mathrm{M}})$, while effect on the selected bandwidth due to the difference between $V(h_{\mathrm{M}})$ and $V_{\mathrm{A}}(h_{\mathrm{M}})$ is asymptotically negligible. In fact, under suitable regularity conditions, we have $B(h_{\mathrm{M}}) - B_{\mathrm{A}}(h_{\mathrm{M}}) = O(h_{\mathrm{M}}^{(2r+2)})$, which determines that the relative difference between h_{M} and h_{A} is of order $O(h_{\mathrm{M}}^2) = O(n^{-2/(2r+1)})$. However, it can be easily shown that $V(h_{\mathrm{M}}) - V_{\mathrm{A}}(h_{\mathrm{M}}) = O[(nh_{\mathrm{M}})^{-2}] = O(h_{\mathrm{M}}^{4r})$. The change in the selected bandwidth caused by using $V_{\mathrm{A}}(h)$ as an approximation of $V(h)$ is of the relative order $O_p(h_{\mathrm{M}}^{2r}) = O_p(n^{-2r/(2r+1)}) = o_p(n^{-1/2})$ and is hence asymptotically negligible for any bandwidth selection rule.

Hence, for the ordinary and modified DS bandwidth selectors we have

Proposition 1. *Under the same conditions, the MDS bandwidth selector has the same asymptotic properties as the ordinary one up to an $o_p(n^{-1/2})$ term.*

The practical performance of the ordinary and modified DS bandwidth selectors will be compared in section 4 through simulation.

2.4 An iterative double smoothing procedure

Like the PL method, a DS bandwidth selector is data-driven, only then if the pilot bandwidth g is also selected based on the data. This seems to be a paradoxical. In the proposal in Feng and Heiler (1999), denoted by $\hat{h}_{\mathrm{ODS}}$, the bandwidth $\hat{g}_{\mathrm{RC}}$ selected following the R-criterion (Rice, 1984) is used in the pilot estimate. However, the use of $\hat{g}_{\mathrm{RC}}$ has two disadvantages: 1. $\hat{h}_{\mathrm{ODS}}$ shares in part the disadvantage of $\hat{g}_{\mathrm{RC}}$ and hence has large finite sample variation; 2. Like $V(h)$, the R-criterion depends strongly on the iid assumption, and it is difficult to extend it to nonparametric regression with dependent data.

In the following an IDS procedure will be proposed without using other methods for bandwidth selection. The name IDS shows that this proposal follows the IPL idea of Gasser et al. (1991) and is also based on fixpoint search. Let an r-th order kernel K and an s-th order kernel L be used in the main and the pilot stages, respectively. And let $1/(2r+1) << \beta << 1$ and $0 < \alpha < 1/(2r+1)$. Denote the selected bandwidth by $\hat{h}_{\mathrm{IDS}}$. Then the IDS algorithm is defined as follows

1. Set $g_0 = n^{-\beta}$ and set $j = 1$;

2. In the j-th iteration set $g_j = \hat{h}_{j-1} n^{\alpha}$;

3. Select $\hat{h}_j$ by minimizing $\hat{M}(h, g_j)$ or $\hat{M}_{\mathrm{M}}(h, g_j)$, respectively;

4. Stop the procedure, when $\hat{h}_j$ converges or until a given maximal number (N) of iterations and set $\hat{h}_{\mathrm{IDS}} = \hat{h}_j$, otherwise increase j by 1 and go back to step 2.

Here $g_0 = n^{-\beta}$ is called the starting pilot bandwidth and $\alpha > 0$ the inflation factor. It will be shown that, for any $0 < \beta < 1$ and $0 < \alpha < 1/(2r+1)$, $\hat{h}_{\mathrm{IDS}}$ is a bandwidth selector with given rate of convergence, which only depends on α. Following Heiler and Feng (1998), the optimal choice of α is

$$\alpha = \frac{1}{2r+1} - \frac{1}{2r+s+1} = \frac{s}{(2r+1)(2r+s+1)}. \tag{2.11}$$

However, there is no objective method for choosing g_0. A large g_0 will reduce the required number of iterations. Asymptotically, if g_0 is chosen such that $g_1/h_{\mathrm{M}} \to \infty$, then $\hat{h}_1$ will be asymptotically optimal. However, g_0 should not be too large and at the same time it should also not be too small, since a too large g_0 may cause oversmoothing, whereas a too small g_0 may introduce the danger of undersmoothing. In this paper the use of $\beta \simeq 1/2$ is proposed.

2.5 The main proposal

Although the iterative idea can be directly used for selecting the bandwidth h, in this paper we would like to use it only as a pilot method of another DS bandwidth selector in order to reduce the effect of the subjectively chosen parameter β on the final selected bandwidth. The possibility of using the IDS procedure directly will be investigated elsewhere. For simplicity, the following bandwidth selector will be proposed for $r = 2$ and $s = 4$ only. At the pilot stage of the IDS procedure, a 4-th order kernel L_p with a bandwidth g_p will be used, as well, so that the highest kernel order required is equal to 4.

For the pilot IDS procedure we have $r = r_p = 4$, $s = s_p = 4$ and $\alpha = 4/117$, where $\beta = 61/117$ and $N = 15$ are used. Our main proposal, $\hat{h}_{\mathrm{MDS}}$, is as follows

1. Select the pilot bandwidth $\hat{g}_{\mathrm{IDS}}$:

   a) Set $g_{p0} = n^{-61/117}$ and set $j = 1$;

   b) In the j-th iteration set $g_{pj} = \hat{g}_{j-1} n^{4/117}$;

   c) Select $\hat{g}_j$ by minimizing $\hat{M}_{\mathrm{M}}(g, g_{pj})$;

d) Stop the procedure, when $\hat{g}_j$ converges or at the 15-th iteration and set $\hat{g}_{\mathrm{IDS}} = \hat{g}_j$, otherwise increasing j by 1 and go back to step b).

2. Select $\hat{h}$ by minimizing $\hat{M}_{\mathrm{M}}(h, \hat{g}_{\mathrm{IDS}})$;

Remark 1. Here $\beta = 61/117 \simeq 0.5$ is chosen so that $-\beta + 12 * \alpha = 13/117 = 1/9$. Now, $\hat{g}_{\mathrm{IDS}}$ is of order $n^{-1/9}$ after at most 12 iterations and it is optimal after at most 13 iterations (see Corollary 1 in the next section). As in Gasser et al. (1991), we propose further two iterations to improve the finite sample property of $\hat{g}_{\mathrm{IDS}}$. Note, however, that N is just the maximal number of required iterations. And the procedure will often converge before N iterations have been done.

3 Asymptotic results

In this section the asymptotic properties of a general IDS bandwidth selector $\hat{h}_{\mathrm{IDS}}$ will be discussed at first. Then the asymptotic properties of $\hat{h}_{\mathrm{MDS}}$ are investigated and compared with those of $\hat{h}_{\mathrm{ODS}}$ and $\hat{h}_{\mathrm{IPL}}$.

3.1 Results on $\hat{h}_{\mathrm{IDS}}$

It is assumed that the bandwidth h satisfies $h \to 0$, $nh \to \infty$ as $n \to \infty$. Similar conditions on the pilot bandwidths g and g_p are also assumed. Further assumptions are

A1. K and L are compactly supported, $K^{(s+1)}$ and $L^{(r+1)}$ are bounded.

A2. Assume that $m^{(r+s)}$ is continuous on (0, 1).

A3. Assume that $E(\epsilon^4) < \infty$ and that $\hat{\sigma}^2$ as defined in (2.1) is used.

For our main results only K' (not $K^{(s+1)}$) in A1 has to be bounded (see Härdle et al., 1992 and Heiler and Feng, 1998). Denote by $\hat{h}_{\mathrm{DS}}$ the bandwidth selected by a general DS procedure. In order that $\hat{h}_{\mathrm{DS}}$ is optimal, the relationship $h_{\mathrm{M}}/g \to 0$ as $n \to \infty$ has to be fulfilled (see Müller, 1985 and Heiler and Feng, 1998). The following proposition gives details on the behaviour of $\hat{h}_{\mathrm{DS}}$ corresponding to the relationship between g and h_{M}.

Proposition 2. *Under the assumptions A1. to A3., the following holds for* $\hat{h}_{\mathrm{DS}}$*:*

i) If $g = o(h_{\mathrm{M}})$*, then* $\hat{h}_{\mathrm{DS}}$ *is at least of the order* $O_p(g)$*;*

ii) If $g = O(h_{\mathrm{M}})$*, then* $\hat{h}_{\mathrm{DS}} = O_p(h_{\mathrm{M}})$*, but is not yet asymptotically optimal;*

iii) If $h_{\mathrm{M}} = o(g)$*, then* $\hat{h}_{\mathrm{DS}} = h_{\mathrm{M}}(1+o_p(1))$*, i.e.* $\hat{h}_{\mathrm{DS}}$ *is now asymptotically optimal.*

The proof of proposition 2 is given in the appendix.

Remark 2. Proposition 2 gives some insights on the DS idea and is the basis for the development of the IDS method. For given r, s and β, the required maximal number of iterations for $\hat{h}_{\mathrm{IDS}}$ can be calculated following this proposition. In Case 1 of Proposition 2, just a lower bound for the selected bandwidth is given, since here the exact order is random (not fixed).

Remark 3. Proposition 2 shows that, if $\alpha > 0$, then $\hat{h}_{\mathrm{IDS}}$ will be optimal after some iterations and is always optimal afterwards.

The asymptotic properties of an IDS bandwidth selector are the same as those of a common DS bandwidth selector with a pilot bandwidth $g = h_{\mathrm{M}} n^{\alpha}$, which are quantified by Theorem 1 in Heiler and Feng (1998). Let λ_s denote the kernel constant of L and c_0 the constant defined in (2.8). Let c_1 and c_2 be the two constants such that

$$M''(h_{\mathrm{M}}) \doteq c_1(nh_{\mathrm{M}}^3)^{-1} \doteq c_2 h_{\mathrm{M}}^{2r-2}. \tag{3.1}$$

Let α is as defined in (2.11) and denote by N^0 the maximal number of iterations, so that $\hat{h}_{\mathrm{IDS}}$ is of order $O_p(n^{-1/(2r+1)})$. Then, following Heiler and Feng (1998), we have

Theorem 1: *Under the assumptions A1. to A3. We have, after at most* $N^0 + 3$ *iterations,*

$$\begin{aligned}(\hat{h}_{\mathrm{IDS}} - h_{\mathrm{M}})/h_{\mathrm{M}} \;&=\; \gamma_1(\hat{\sigma}^2 - \sigma^2) + (\gamma_2 c_0^{-(4r+1)} n^{-(2s+1)/(2r+s+1)} + \gamma_3 n^{-1})^{1/2} Z_n \\ &+\; [\gamma_4 c_0^s + \gamma_5 c_0^{-(2r+1)}] n^{-s/(2r+s+1)}(1 + o(1))\,,\end{aligned} \tag{3.2}$$

where Z_n is asymptotically normal $N(0,1)$, the γ_1, ..., γ_5 are constants given by

$$\begin{aligned}
\gamma_1 &= c_1^{-1}(d-c)\int K^2(y)dy, \\
\gamma_2 &= 4c_2^{-2}r^2(d-c)\kappa_r^4\sigma^4 \int \left[\int L^{(r)}(y)L^{(r)}(y+z)dy\right]^2 dz, \\
\gamma_3 &= 16c_2^{-2}r^2\kappa_r^4\sigma^2 \int_c^d (m^{(2r)}(x))^2 f(x)dx, \\
\gamma_4 &= -4c_2^{-1}r\kappa_r^2\lambda_s \int_c^d m^{(r)}(x)m^{(r+s)}(x)f(x)dx, \qquad \textit{and} \\
\gamma_5 &= -2c_2^{-1}r(d-c)\sigma^2\kappa_r^2 \int (L^{(r)})^2.
\end{aligned}$$

The proof of Theorem 1 is omitted. The rate of convergence of $\hat{h}_{\mathrm{IDS}}$ with α as defined in (2.11) is $n^{-\frac{s}{2r+s+1}}$ if $s \leq 2r$ or $n^{-\frac{1}{2}}$ if $s \geq 2r+2$. It is $n^{-4/9}$ e.g. for $r=2$ and $s=4$.

Let g_{M} denote the optimal bandwidth and c_s the constant in (2.8) for a kernel estimate with the s-th order kernel L. Then the rate of convergence of $\hat{g}_{\mathrm{IDS}}$ is $n^{-4/13}$ after at most $N^0+3=15$ iterations, where $N^0=12$.

Corollary 1: *Under similar assumptions as A1. to A3. We have, after at most 15 iterations,*

$$\begin{aligned}
(\hat{g}_{\mathrm{IDS}} - g_{\mathrm{M}})/g_{\mathrm{M}} &= \gamma_1(\hat{\sigma}^2-\sigma^2) + (\gamma_2 c_s^{-17} n^{-9/13} + \gamma_3 n^{-1})^{1/2} Z_n \\
&+ [\gamma_4 c_s^4 + \gamma_5 c_s^{-9}] n^{-4/13}(1+o(1))\,,
\end{aligned} \tag{3.3}$$

where Z_n is as before, c_s denotes the constant in (2.8) defined for L, and γ_1, ... ,γ_5 are as defined in Theorem 1 with $r=4$, $s=4$ and corresponding adaptation to the kernel functions.

Note that the assumptions for Corollary 1 have also to adapted to the kernel function used in the pilot and main stages. The proof of this corollary is omitted.

3.2 Results on $\hat{h}_{\mathrm{MDS}}$

Note that, the use of $\hat{g}_{\mathrm{IDS}}$ as a pilot bandwidth for selecting h is the optimal choice up to a constant (see Heiler and Feng, 1998). Hence, $\hat{h}_{\mathrm{MDS}}$ has the highest rate of convergence in the case with $r=2$ and $s=4$. Define $m_3 = E(\epsilon^3)$ and $m_4 = E(\epsilon^4)$. The following theorem gives more detailed results on $\hat{h}_{\mathrm{MDS}}$. In order to compare these results with those on $\hat{h}_{\mathrm{IPL}}$, results in *ii*) and *iii*) of this theorem are represented in a similar way as in Herrmann (1994).

Theorem 2. *Under the assumptions A1. to A3. We have, for* $\hat{h}_{\mathrm{MDS}}$,

i)

$$\begin{aligned}(\hat{h}_{\mathrm{MDS}} - h_{\mathrm{M}})/h_{\mathrm{M}} &= \gamma_1(\hat{\sigma}^2 - \sigma^2) + (\gamma_2 c_0^{-9} + \gamma_3)^{1/2} n^{-1/2} Z_n \\ &+ [\gamma_4 c_0^4 + \gamma_5 c_0^{-5}] n^{-4/9} (1 + o(1)) ,\end{aligned} \tag{3.4}$$

where Z_n *is asymptotically normal* $N(0,1)$, γ_1, ... ,γ_5 *are as defined in Theorem 1 with* $r = 2$, $s = 4$.

ii)

$$n^{7/10}(\hat{h}_{\mathrm{MDS}} - h_{\mathrm{M}} - O(n^{-29/45})) \longrightarrow N(0, \sigma_2^2), \tag{3.5}$$

in distribution, where

$$\sigma_2^2 = \frac{c_0^2}{25}\left[\frac{35}{9} + (\frac{m_4}{\sigma^4} - 3) + 2\frac{m_3}{\sigma^2}\frac{\int_c^d m^{(4)}}{I(m'')} + 4\sigma^2 \frac{\int_c^d \{m^{(4)}\}^2}{I(m'')^2}\right] + c_0^{-7}\gamma_2. \tag{3.6}$$

iii)

$$cov(n^{7/10}(\hat{h}_{\mathrm{MDS}} - h_{\mathrm{M}} - O(n^{-29/45})), n^{3/10} h_{\mathrm{ASE}}) = \sigma_{12}, \tag{3.7}$$

where

$$\sigma_{12} = -\frac{2}{25\kappa_2}\left[\frac{m_3 \int_c^d m''}{\sigma^2 I(m'')} + 2\frac{\sigma^2 \int_c^d \{m^{(4)}\}^2}{I(m'')^2}\right]. \tag{3.8}$$

If $m_3 = 0$, then $\sigma_{12} < 0$, i.e., like most existing bandwidth selectors, $\hat{h}_{\mathrm{MDS}}$ is asymptotically negatively correlated with h_{ASE} for symmetrically distributed errors. Theorem 2 allows us to compare the asymptotic properties of $\hat{h}_{\mathrm{MDS}}$ with those of $\hat{h}_{\mathrm{IPL}}$. Some differences between the asymptotic properties of $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{IPL}}$ are:

1. The dominating bias term of $\hat{h}_{\mathrm{MDS}}$ is caused by the bias in the pilot smoothing and is of the relative order $n^{-4/9}$, while the bias term of $\hat{h}_{\mathrm{IPL}}$ is due to the approximation in h_{A} and is of the relative order $n^{-1/5}$.

2. The asymptotic variances of both bandwidth selectors are of the same, highest relative order $n^{-1/2}$. By comparing σ_2^2 in (3.6) with those given in (6) in Herrmann (1994) we can see that the constant of the asymptotic variance of $\hat{h}_{\mathrm{MDS}}$ is larger than that of $\hat{h}_{\mathrm{IPL}}$ with the additional term $c_0^{-7}\gamma_2 > 0$. Hence, $\hat{h}_{\mathrm{IPL}}$ is more stable than $\hat{h}_{\mathrm{MDS}}$ but with a larger bias and slower rate of convergence.

3. For symmetrically distributed errors, i.e. with $m_3 = 0$, both $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{IPL}}$ are asymptotically negatively correlated with h_{ASE}. Note that the asymptotic covariance between these two bandwidth selectors and h_{ASE} is the same. Hence, the asymptotic correlation coefficient between $\hat{h}_{\mathrm{MDS}}$ and h_{ASE} is smaller than the one between $\hat{h}_{\mathrm{IPL}}$ and h_{ASE}.

What is the difference between the asymptotic performances of $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{ODS}}$? Both have the same asymptotic properties w.r. to the first term. They differ only in a second term, which is asymptotically negligible. However, $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{ODS}}$ perform quite differently for finite samples, since the rates of convergence of their pilot bandwidths are quite different, namely $O(n^{-4/13})$ for $\hat{g}_{\mathrm{IDS}}$ and $O_p(n^{-1/18})$ for $\hat{g}_{\mathrm{RC}}$ respectively. The variance term of $\hat{g}_{\mathrm{IDS}}$ converges slightly a little faster. Note that the variance of the final selected bandwidth depends strongly on the variance of the pilot bandwidth. It is expected that the finite sample variation in $\hat{h}_{\mathrm{MDS}}$ should be much smaller than that in $\hat{h}_{\mathrm{ODS}}$.

Furthermore, all bandwidth selectors, $\hat{h}_{\mathrm{MDS}}$, $\hat{h}_{\mathrm{ODS}}$ and $\hat{h}_{\mathrm{IPL}}$, have the property (1.6), since they come all within $o_p(n^{-3})$ to h_{M}. Hence they are all asymptotically equivalent w.r. to h_{ASE}.

4 Practical performance

A comparative simulation study was carried out to show the practical performances of the bandwidth selectors $\hat{h}_{\mathrm{IPL}}$, $\hat{h}_{\mathrm{ODS}}$ and $\hat{h}_{\mathrm{MDS}}$. Another bandwidth selector, $\hat{h}_{\mathrm{NDS}}$, defined similarly as $\hat{h}_{\mathrm{MDS}}$ but with $\hat{M}_{\mathrm{M}}$ in the procedure being replaced by $\hat{M}$, is included in the simulation in order to show the practical difference of DS bandwidth selectors based on $\hat{M}$ and $\hat{M}_{\mathrm{M}}$, respectively. The asymptotic difference between $\hat{h}_{\mathrm{NDS}}$ and $\hat{h}_{\mathrm{MDS}}$ is very minor. Also included in the simulation is $\hat{h}_{\mathrm{RC}}$ following the R-criterion, which is used as a comparison. The following six regression functions are chosen:

$$
\begin{aligned}
m_0(x) &= 4x, & m_1(x) &= 2\tanh(4(x-0.5)), \\
m_2(x) &= 5.8(\sin(2(x-0.5))^2), & m_3(x) &= 2\sin(2(x-0.5)\pi), \\
m_4(x) &= 2x + 3\exp(-100(x-0.5)^2), & m_5(x) &= 2\sin(6(x-0.5)\pi),
\end{aligned}
$$

where $x \in [0,1]$. The range of all of these functions is about 4. Standard iid normally distributed errors are used for all regression functions. These regression functions are chosen, because they are quite different with respect to "complicity", and hence have quite different optimal bandwidths for errors with the same distribution (see Figure 1). Note that $\hat{h}_{\mathrm{IPL}}$ is not asymptotically optimal for m_0 but the others are.

The simulation was carried out for n =50, 100, 200 and 400. 400 replications have been carried out for each case. The Epanechnikov kernel (the optimal second order kernel) was used for calculating $\hat{m}$. The second derivative m'' for $\hat{h}_{\mathrm{IPL}}$ is estimated by the corresponding optimal kernel (see Müller, 1988). In the pilot smoothing of a DS bandwidth selector an optimal kernel of order 4 was used. In the pilot stage of the pilot smoothing for $\hat{h}_{\mathrm{NDS}}$ and $\hat{h}_{\mathrm{MDS}}$ a fourth order kernel with degrees of smoothness 3 (Müller, 1988) were used. In this simulation only bandwidths h or g, respectively, such that nh or ng is an integer are considered. All of the selected bandwidths, except for $\hat{h}_{\mathrm{IPL}}$, are obtained by a search based on an optimizing procedure on the range from r/n or s/n, respectively, to $0.5 - 1/n$ (the largest allowed bandwidth).

Box-plots of the 400 replications for the five bandwidth selectors as well for h_{ASE} are shown in Figures 2 through 5. Some detailed statistics on the simulation results are given in Tables 1 to 4, where the first two rows are the true values of h_{M} and $M(h_{h_{\mathrm{M}}})$. Other statistics are the mean, the standard deviation (SD) for each bandwidth selector and for h_{ASE}. Also given in these tables are standard deviation from h_{ASE} (SDO) for each bandwidth selector and for h_{M}, as well as the means of ASE of the estimated regression function in 400 replications ($\overline{\mathrm{ASE}}$) for each bandwidth selector, h_{ASE} and h_{M}.

In the following, the bandwidth selectors will be assessed at first following $\overline{\mathrm{ASE}}$. Note that this is asymptotically equivalent to the assessment following SDO, i.e. by taking h_{ASE} to be the optimal bandwidth (see Hall and Johnstone, 1992). To this end the ratio (%) between the mean of $\overline{\mathrm{ASE}}(h_{\mathrm{ASE}})$ and that for a bandwidth selector will be used (see Table 5), which will be called the empirical efficiency of a bandwidth selector. Note however that, here 100% is not achievable, no matter how large n is. From Table 5 we see that, while the three double smoothing bandwidth selectors have the same asymptotic properties, $\hat{h}_{\mathrm{NDS}}$ and $\hat{h}_{\mathrm{MDS}}$ perform in general much better than $\hat{h}_{\mathrm{ODS}}$. The practical performances of $\hat{h}_{\mathrm{NDS}}$ and $\hat{h}_{\mathrm{MDS}}$ are quite similar. This means that the finite sample performance will not be clearly changed by using $\hat{M}_{\mathrm{M}}$ instead of $\hat{M}$ (for this reason, discussion on the performance of $\hat{h}_{\mathrm{NDS}}$ will be ignored in the following). For the three regression functions m_3, m_4 and m_5, $\hat{h}_{\mathrm{IPL}}$ performs sometimes slightly better than $\hat{h}_{\mathrm{MDS}}$. But the difference between their practical performance is not clear, especially when n is large. For the three regression functions m_0, m_1 and m_2, $\hat{h}_{\mathrm{MDS}}$ performs clearly better than $\hat{h}_{\mathrm{IPL}}$. Although $\hat{h}_{\mathrm{IPL}}$

and $\hat{h}_{\mathrm{ODS}}$ perform quite differently, their practical performances are comparable on average.As expected, $\hat{h}_{\mathrm{RC}}$ performs in all cases the worst, except for m_1 with $n = 400$, where it performs slightly better than $\hat{h}_{\mathrm{IPL}}$. The assessment following $\overline{\mathrm{ASE}}$ gives evidence for choosing $\hat{h}_{\mathrm{MDS}}$.

The practical performance of a bandwidth selector can also be assessed following the distance to h_{M}. Some changes in this case are: Firstly, the differences between the selected methods following this criterion are much larger than those following $\overline{\mathrm{ASE}}$; Secondly, following this criterion, $\hat{h}_{\mathrm{ODS}}$ performs on the average better than $\hat{h}_{\mathrm{IPL}}$ for large n; Thirdly, following this criterion, even $\hat{h}_{\mathrm{RC}}$ performs better than $\hat{h}_{\mathrm{IPL}}$ in the case of m_0 with all n's, since $\hat{h}_{\mathrm{IPL}}$ is now not asymptotically optimal but $\hat{h}_{\mathrm{RC}}$ is. Furthermore, we can find that, the improvement in $\hat{h}_{\mathrm{MDS}}$ in comparison with $\hat{h}_{\mathrm{ODS}}$ is mainly due to the reduction in variance. Moreover, $\hat{h}_{\mathrm{IPL}}$ has the smallest variances in almost all cases. This means that $\hat{h}_{\mathrm{IPL}}$ is the most stable method. Its bad performance in some cases is due to the unacceptably large bias. All of the simulation results confirm the theoretical findings. Note in particular that, in most of the cases $\hat{h}_{\mathrm{MDS}}$, and in some of the cases also $\hat{h}_{\mathrm{ODS}}$ and $\hat{h}_{\mathrm{IPL}}$ perform even better than h_{ASE}, since they all have a higher rate of convergence to h_{M} than h_{ASE}. Now, the evidence for choosing $\hat{h}_{\mathrm{MDS}}$ is stronger. In the extreme case of m_2 with $n = 400$, we find that $\hat{h}_{\mathrm{MDS}}$ is nearer to h_{M} than $\hat{h}_{\mathrm{IPL}}$ in all of the 400 replications (see Figure 6). But this is not true following the distance to h_{ASE}.

By comparing the accuracy of the selected bandwidth and of $\hat{m}$ over all regression functions we can find that, if m is easy to estimate, i.e. when the structure of the regression function is relatively simple, then the bandwidth is difficult to select, and vice versa. This seems to be a paradoxical. However, it can be reasonably explained, e.g. for the first case. On one hand, h_{M} is large in this case. A simple representation of (3.2) or of (3.4) shows that, in general, the larger h_{M}, the larger the (asymptotic) variance of a bandwidth selector. On the other hand, $\hat{m}$ is now not so sensitive to the change in the selected bandwidth. The accuracy of $\hat{m}$ is quite similar for a wide range of bandwidths. And hence, in this case, bandwidth selection plays a relatively unimportant role. A similar phenomenon was reported by Härdle et al. (1988), where the accuracy of the selected bandwidth and of the kernel estimators with kernels of different orders are considered.

The practical performance of a bandwidth selector can also be investigated by considering the correlation coefficient with h_{ASE}. At first sight, the larger h_{ASE} is, the larger the selected bandwidth should be. When this is so, then the bandwidth selector will have a positive correlation with h_{ASE}. Unfortunately, most of the proposed bandwidth selectors have a negative correlation with h_{ASE} as mentioned e.g. in Härdle et al. (1988) and Herrmann (1994) (see however Hall and Johnstone, 1992 for an exception). Correlation coefficients for all of these bandwidth selectors calculated from each of 400 replications are reported in Table 6. We see that, they are always clearly negative. Moreover, we find another seeming paradoxical phenomenon, which is also reported by Härdle et al. (1988), namely a better bandwidth selector seems to have a stronger negative correlation! By looking at the simulation results more exactly, we can see that this is simply due to the fact that a better bandwidth selector has in general a smaller variance. The negative correlation between the bandwidth selectors and h_{ASE} is shown in Figures 6 and 7, where the bandwidths selected in the 400 replications are shown against h_{ASE} for the case of m_2 and m_3 with $n = 400$. Figures 6 and 7 also give us some insight about the practical performance of the selected methods, for instance, how the bad performance of $\hat{h}_{\mathrm{RC}}$ is improved by $\hat{h}_{\mathrm{ODS}}$ and then by $\hat{h}_{\mathrm{MDS}}$, and what the advantages and disadvantages of $\hat{h}_{\mathrm{IPL}}$ are compared with $\hat{h}_{\mathrm{MDS}}$.

5 Concluding remarks

In this paper, a modified DS bandwidth selector $\hat{h}_{\mathrm{MDS}}$ is proposed with an IDS procedure at the pilot stage. It is shown, theoretically and by simulations, that the DS idea should be used as the standard approach for bandwidth selection in nonparametric regression. Some further arguments that support this conclusion are: 1. The DS rule can easily be adapted to bandwidth selection in nonparametric decomposition of seasonal time series (see Heiler and Feng, 2000), whereas the plug-in method is not suitable; 2. The DS idea makes it possible to select the bandwidth for each component separately in a model with unknown components, such as the time series decomposition model mentioned above.

To our knowledge, this is the first comparative study between the DS and the plug-in ideas. Our study also shows that $\hat{h}_{\mathrm{IPL}}$ (Gasser et al., 1991) has some advantages. Firstly, the procedure of $\hat{h}_{\mathrm{IPL}}$ is much simpler than the one for $\hat{h}_{\mathrm{MDS}}$ and the computing time for $\hat{h}_{\mathrm{IPL}}$ is practically negligible in comparison with that for $\hat{h}_{\mathrm{MDS}}$. Secondly, the order of existing continuous derivatives required for the asymptotic results is 4 for $\hat{h}_{\mathrm{IPL}}$, while it is 8 for $\hat{h}_{\mathrm{MDS}}$. Finally, in many cases, e.g. the cases of m_3, m_4 and m_5, the practical performance of $\hat{h}_{\mathrm{IPL}}$ is not worse than that of $\hat{h}_{\mathrm{MDS}}$ for small or moderate n. Hence, $\hat{h}_{\mathrm{IPL}}$ is still one of the best methods for bandwidth selection in nonparametric regression.

References

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

[2] Beran, J. and Feng, Y. (1999). Locally polynomial fitting with long-range dependent errors. Discussion paper No. 99/07, Center of Finance and Econometrics, University of Konstanz.

[3] Beran, J. and Feng, Y. (2000). Data-driven estimation of semiparametric fractional autoregressive models. Discussion paper No. 00/16, Center of Finance and Econometrics, University of Konstanz.

[4] Chiu, S-T. (1991). Some stabilized bandwidth selectors for nonparametric regression, *Ann. Statist.*, **19**, 1528–1546.

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

[6] Feng, Y. (1999). *Kernel- and Locally Weighted Regression – with Applications to Time Series Decomposition.* Verlag für Wissenschaft und Forschung, Berlin.

[7] Feng, Y. and Heiler, S. (1999). Selecting bandwidth for nonparametric regression based on bootstrap idea. Preprint, University of Konstanz.

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

[9] Gasser, T. and Müller, H.G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, **11**, 171–185.

[10] Gasser, T., Müller, H.G., Köhler, W., Molinari, L. and Prader, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.*, **12**, 210–229.

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

[12] Gasser, T. Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625–633.

[13] 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.

[14] Härdle, W., Hall, P. and Marron, J.S. (1992). Regression smoothing parameters that are not far from their optimum. *J. Amer. Statist. Assoc.*, **87**, 227–233.

[15] Hall, P. and Johnstone, I. (1992). Empirical functionals and efficient smoothing parameter selection. *J. Roy. Statist. Soc. Ser. B*, **54**, 475–530.

[16] Heiler, S. and Feng, Y. (1998). A simple root n bandwidth selector for non-parametric regression. *J. Nonpar. Statist.*, **9**, 1–21.

[17] Heiler, S. and Feng, Y. (2000). Data-driven decomposition of seasonal time series. To apear in *J. Statist. Pl. and Inf.*

[18] Herrmann, E. (1994). Asymptotic distribution of bandwidth selectors in kernel regression estimation. *Statistical Papers*, **35**, 17–26.

[19] Herrmann, E. and Gasser, T. (1994). Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. Preprints, University of Zurich and Damstadt Institute of Technology.

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

[21] Müller, H.-G. (1985). Empirical bandwidth choice for nonparametric kernel regression by means of pilot estimators. *Statist. Decisions*, Supp. Issue **2**, 193–206.

[22] Müller, H.-G. (1988). *Nonparametric Analysis of Longitudinal Data*, Springer-Verlag, Berlin.

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

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

[25] 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.

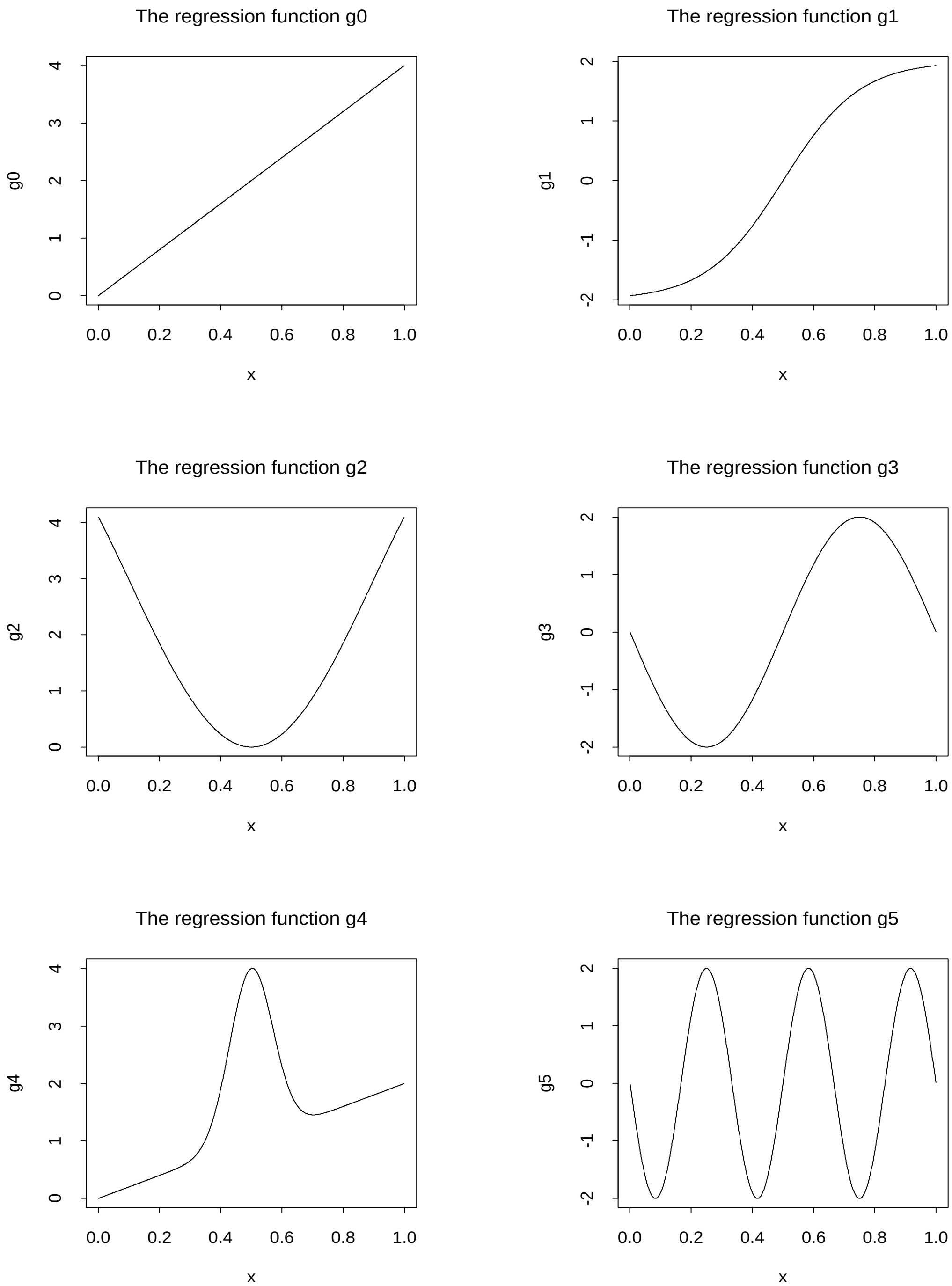

Figure 1: The six regression functions.

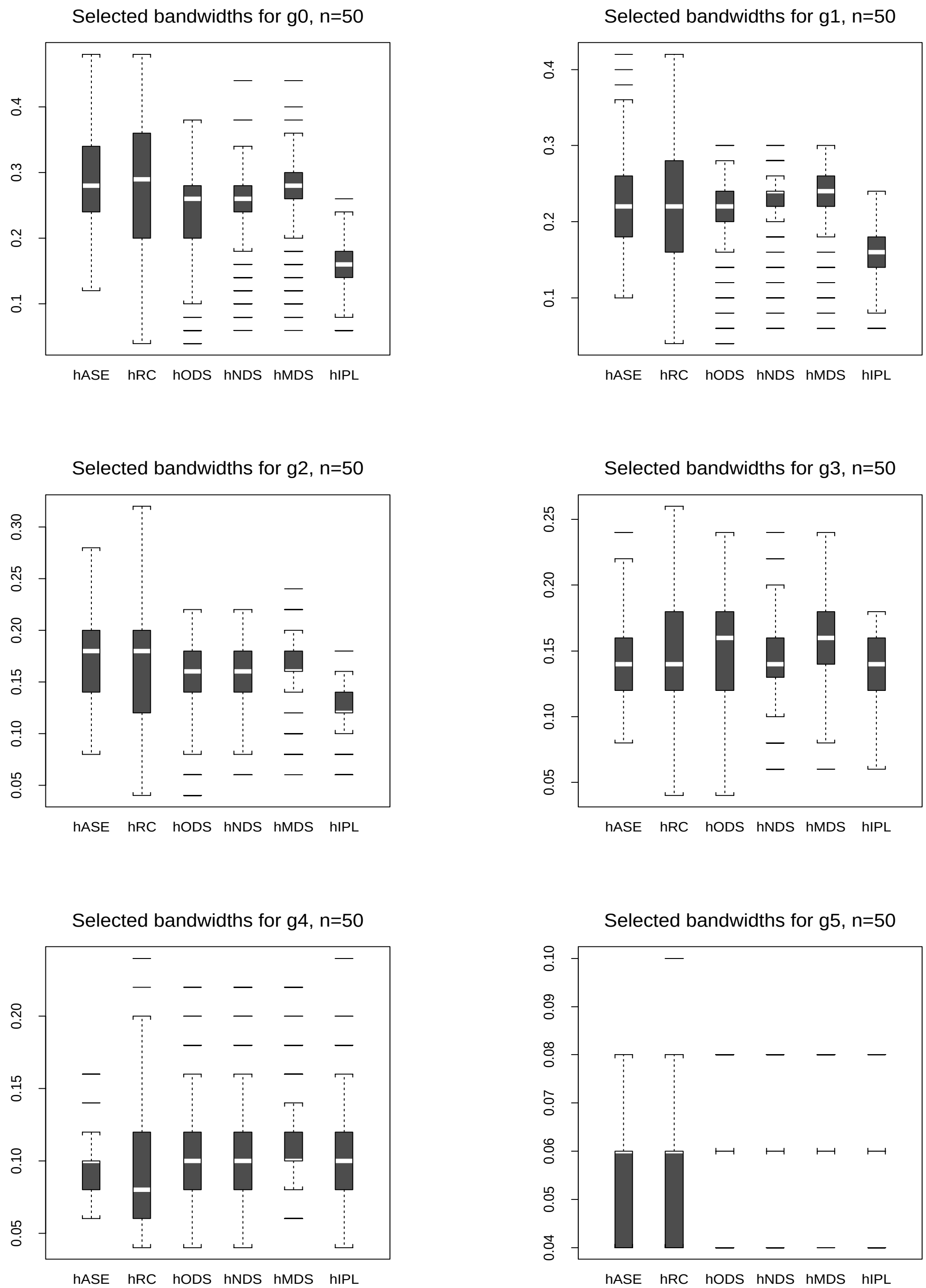

Figure 2: Box-plots for selected bandwidths in all cases with $n = 50$.

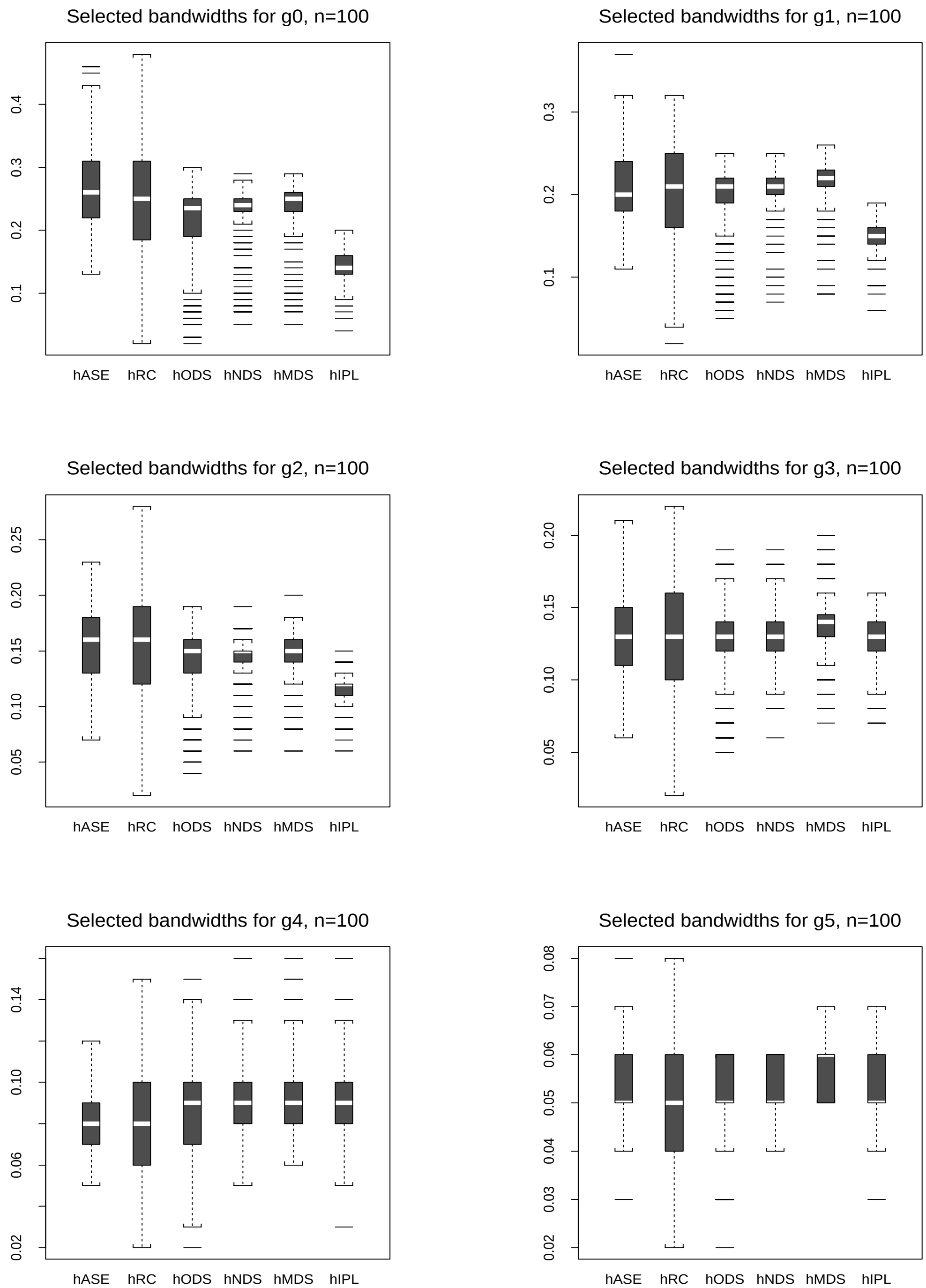

Figure 3: Box-plots for selected bandwidths in all cases with $n = 100$.

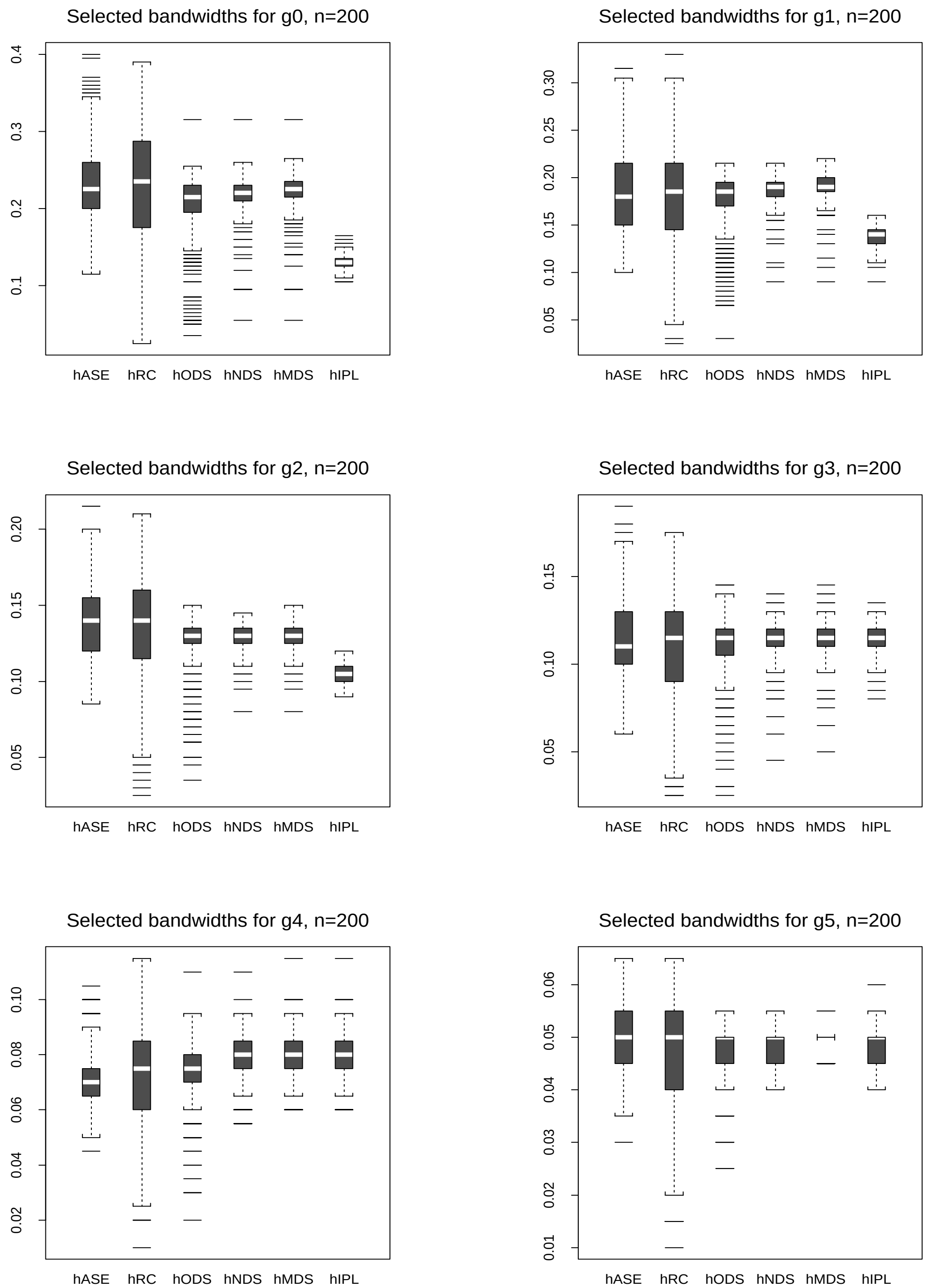

Figure 4: Box-plots for selected bandwidths in all cases with $n = 200$.

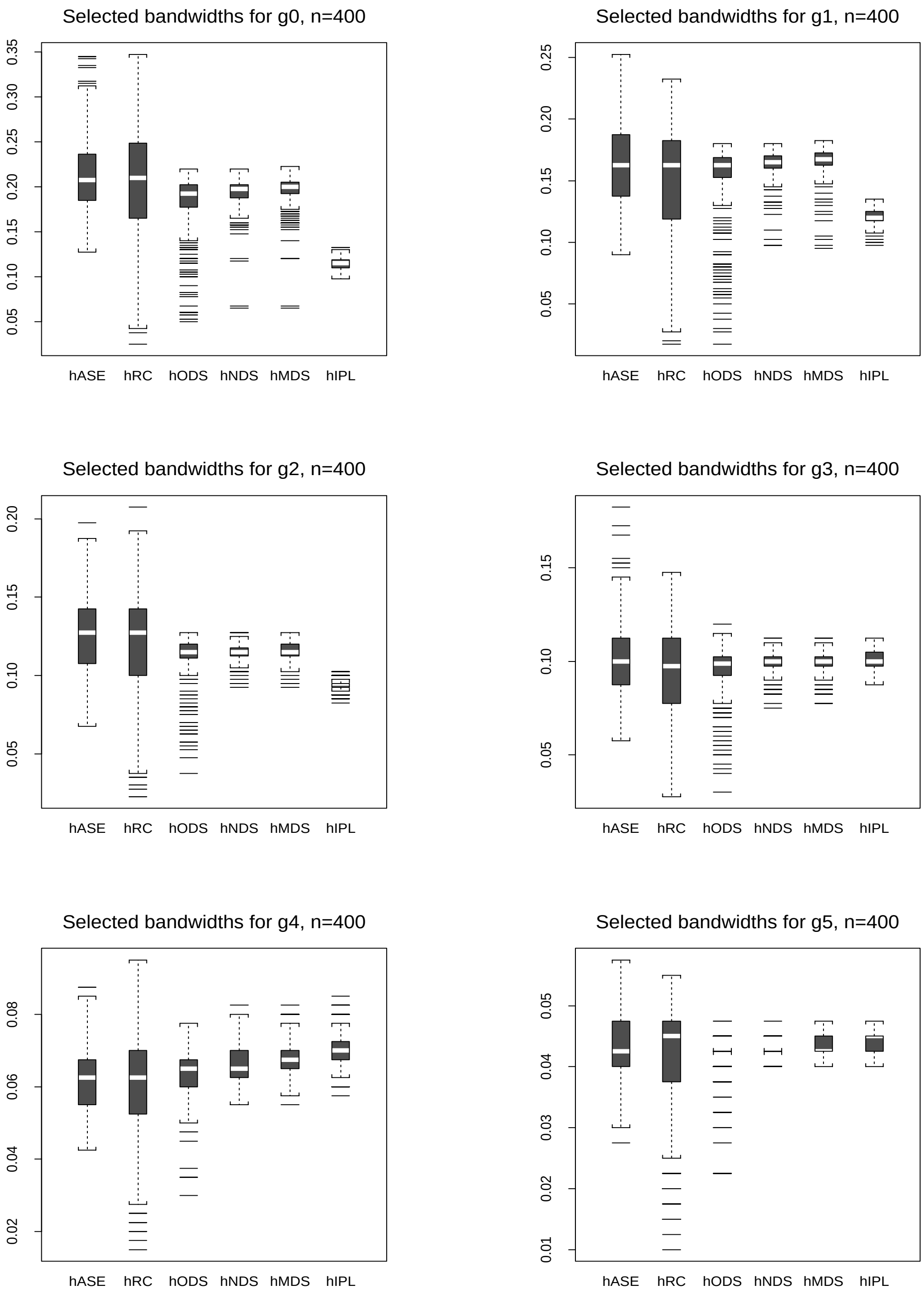

Figure 5: Box-plots for selected bandwidths in all cases with $n = 400$.

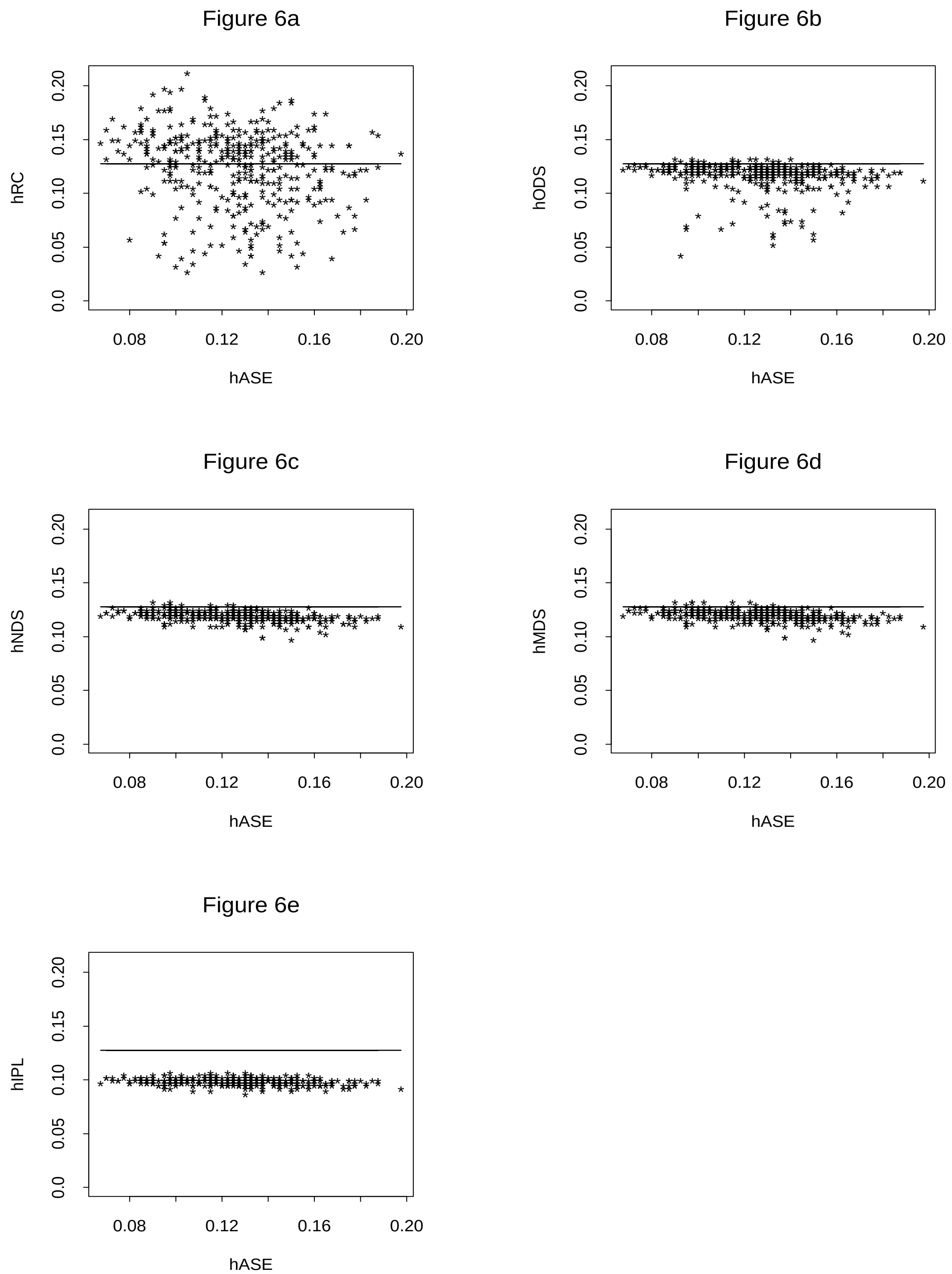

Figure 6: Selected bandwidths against h_{ASE} in 400 replications for the case of m_2 with $n = 400$. Figures 6a trough 6e show the results for $\hat{h}_{\mathrm{RC}}$, $\hat{h}_{\mathrm{ODS}}$, $\hat{h}_{\mathrm{NDS}}$, $\hat{h}_{\mathrm{MDS}}$ and $\hat{h}_{\mathrm{IPL}}$, respectively.

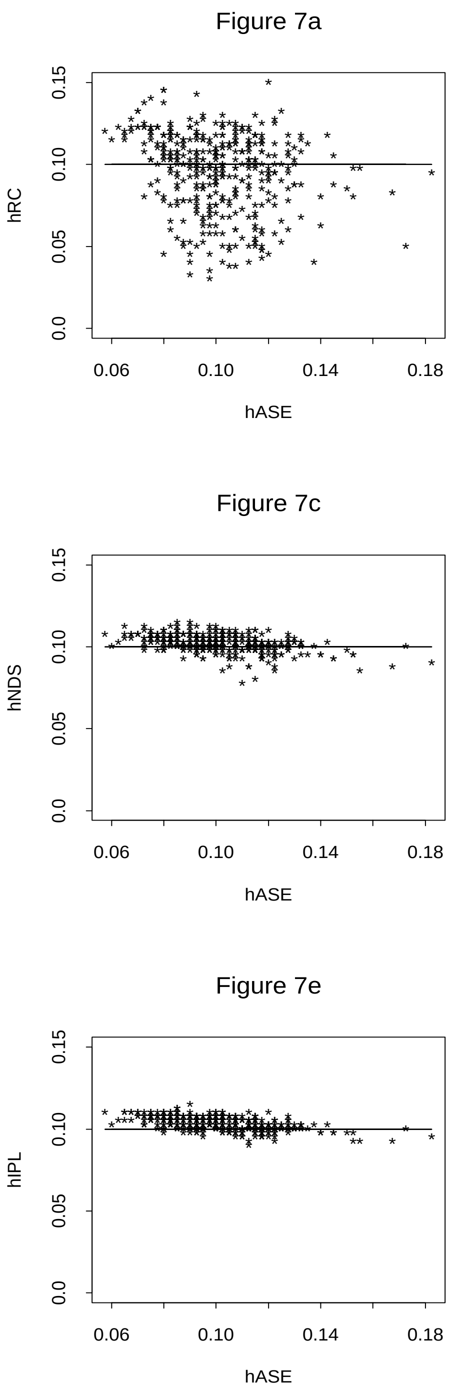

Figure 7: Results as given in Figure 6 but for the case of m_3 with $n = 400$.

Table 1: h_{M}, $\mathrm{M}(h_{\mathrm{M}})$ and statistics from 400 simulation ($n = 50$).

		m_0	m_1	m_2	m_3	m_4	m_5
h_{M}	True	0.3000	0.2400	0.1800	0.1400	0.1000	0.0600
	$\mathrm{M}(h_{\mathrm{M}})$	0.0399	0.0497	0.0639	0.0778	0.1207	0.1785
	$\overline{\mathrm{ASE}}$	0.0388	0.0501	0.0670	0.0766	0.1226	0.1763
h_{ASE}	Mean	0.2902	0.2291	0.1701	0.1481	0.0934	0.0556
	SD	0.0784	0.0561	0.0371	0.0341	0.0178	0.0112
	$\overline{\mathrm{ASE}}$	0.0325	0.0447	0.0613	0.0698	0.1177	0.1720
$\hat{h}_{\mathrm{RC}}$	Mean	0.2794	0.2186	0.1624	0.1422	0.0908	0.0570
	SD	0.1122	0.0832	0.0607	0.0515	0.0383	0.0146
	SDO	0.1568	0.1141	0.0780	0.0703	0.0468	0.0213
	$\overline{\mathrm{ASE}}$	0.0667	0.0777	0.0993	0.1045	0.1546	0.1956
$\hat{h}_{\mathrm{ODS}}$	Mean	0.2341	0.2092	0.1524	0.1465	0.1003	0.0578
	SD	0.0713	0.0552	0.0394	0.0411	0.0323	0.0104
	SDO	0.1319	0.0906	0.0631	0.0604	0.0423	0.0175
	$\overline{\mathrm{ASE}}$	0.0583	0.0662	0.0865	0.0951	0.1441	0.1876
$\hat{h}_{\mathrm{NDS}}$	Mean	0.2483	0.2237	0.1532	0.1483	0.1055	0.0591
	SD	0.0612	0.0434	0.0317	0.0337	0.0301	0.0084
	SDO	0.1222	0.0810	0.0576	0.0557	0.0423	0.0162
	$\overline{\mathrm{ASE}}$	0.0518	0.0591	0.0801	0.0887	0.1412	0.1844
$\hat{h}_{\mathrm{MDS}}$	Mean	0.2636	0.2351	0.1627	0.1581	0.1132	0.0633
	SD	0.0578	0.0414	0.0297	0.0335	0.0304	0.0082
	SDO	0.1155	0.0789	0.0552	0.0565	0.0456	0.0173
	$\overline{\mathrm{ASE}}$	0.0491	0.0576	0.0773	0.0878	0.1428	0.1860
$\hat{h}_{\mathrm{IPL}}$	Mean	0.1556	0.1616	0.1217	0.1353	0.0974	0.0590
	SD	0.0365	0.0347	0.0238	0.0283	0.0291	0.0072
	SDO	0.1634	0.1004	0.0680	0.0533	0.0393	0.0152
	$\overline{\mathrm{ASE}}$	0.0656	0.0677	0.0872	0.0885	0.1398	0.1819

Table 2: h_{M}, $\mathrm{M}(h_{\mathrm{M}})$ and statistics from 400 simulation ($n = 100$).

		m_0	m_1	m_2	m_3	m_4	m_5
h_{M}	True	0.2600	0.2100	0.1600	0.1300	0.0800	0.0500
	$\mathrm{M}(h_{\mathrm{M}})$	0.0219	0.0282	0.0360	0.0448	0.0703	0.1035
	$\overline{\mathrm{ASE}}$	0.0216	0.0260	0.0357	0.0482	0.0692	0.1037
h_{ASE}	Mean	0.2658	0.2071	0.1553	0.1296	0.0821	0.0538
	SD	0.0610	0.0433	0.0309	0.0302	0.0136	0.0078
	$\overline{\mathrm{ASE}}$	0.0186	0.0237	0.0328	0.0439	0.0665	0.1004
$\hat{h}_{\mathrm{RC}}$	Mean	0.2451	0.1994	0.1502	0.1254	0.0801	0.0497
	SD	0.0948	0.0648	0.0496	0.0397	0.0261	0.0133
	SDO	0.1271	0.0870	0.0657	0.0570	0.0340	0.0172
	$\overline{\mathrm{ASE}}$	0.0385	0.0370	0.0530	0.0638	0.0886	0.1200
$\hat{h}_{\mathrm{ODS}}$	Mean	0.2124	0.1928	0.1413	0.1302	0.0864	0.0517
	SD	0.0593	0.0419	0.0278	0.0244	0.0193	0.0078
	SDO	0.1081	0.0675	0.0492	0.0452	0.0290	0.0126
	$\overline{\mathrm{ASE}}$	0.0335	0.0319	0.0435	0.0548	0.0798	0.1096
$\hat{h}_{\mathrm{NDS}}$	Mean	0.2319	0.2093	0.1425	0.1319	0.0919	0.0527
	SD	0.0407	0.0230	0.0185	0.0177	0.0164	0.0050
	SDO	0.0888	0.0545	0.0427	0.0415	0.0281	0.0108
	$\overline{\mathrm{ASE}}$	0.0268	0.0280	0.0403	0.0524	0.0771	0.1060
$\hat{h}_{\mathrm{MDS}}$	Mean	0.2396	0.2155	0.1462	0.1359	0.0950	0.0557
	SD	0.0411	0.0226	0.0184	0.0179	0.0164	0.0051
	SDO	0.0868	0.0548	0.0418	0.0420	0.0294	0.0109
	$\overline{\mathrm{ASE}}$	0.0263	0.0279	0.0398	0.0525	0.0775	0.1061
$\hat{h}_{\mathrm{IPL}}$	Mean	0.1431	0.1526	0.1156	0.1270	0.0906	0.0534
	SD	0.0216	0.0177	0.0118	0.0146	0.0160	0.0056
	SDO	0.1403	0.0748	0.0535	0.0396	0.0274	0.0111
	$\overline{\mathrm{ASE}}$	0.0358	0.0316	0.0432	0.0517	0.0766	0.1065

Table 3: h_{M}, $\mathrm{M}(h_{\mathrm{M}})$ and statistics from 400 simulation ($n = 200$).

		m_0	m_1	m_2	m_3	m_4	m_5
h_{M}	True	0.2350	0.1850	0.1400	0.1150	0.0700	0.0500
	$\mathrm{M}(h_{\mathrm{M}})$	0.0120	0.0160	0.0204	0.0258	0.0409	0.0595
	$\overline{\mathrm{ASE}}$	0.0116	0.0167	0.0205	0.0258	0.0406	0.0574
h_{ASE}	Mean	0.2327	0.1848	0.1408	0.1141	0.0716	0.0481
	SD	0.0479	0.0420	0.0255	0.0232	0.0102	0.0064
	$\overline{\mathrm{ASE}}$	0.0104	0.0151	0.0190	0.0238	0.0392	0.0556
$\hat{h}_{\mathrm{RC}}$	Mean	0.2278	0.1782	0.1353	0.1080	0.0705	0.0464
	SD	0.0772	0.0533	0.0358	0.0304	0.0178	0.0106
	SDO	0.0976	0.0787	0.0490	0.0432	0.0235	0.0138
	$\overline{\mathrm{ASE}}$	0.0179	0.0226	0.0263	0.0329	0.0480	0.0663
$\hat{h}_{\mathrm{ODS}}$	Mean	0.2029	0.1736	0.1266	0.1115	0.0744	0.0475
	SD	0.0425	0.0331	0.0188	0.0180	0.0113	0.0047
	SDO	0.0742	0.0624	0.0376	0.0339	0.0187	0.0093
	$\overline{\mathrm{ASE}}$	0.0149	0.0199	0.0233	0.0291	0.0443	0.0596
$\hat{h}_{\mathrm{NDS}}$	Mean	0.2172	0.1874	0.1296	0.1142	0.0778	0.0482
	SD	0.0213	0.0142	0.0079	0.0103	0.0078	0.0027
	SDO	0.0602	0.0509	0.0318	0.0291	0.0173	0.0081
	$\overline{\mathrm{ASE}}$	0.0126	0.0176	0.0215	0.0272	0.0429	0.0583
$\hat{h}_{\mathrm{MDS}}$	Mean	0.2220	0.1912	0.1312	0.1158	0.0790	0.0494
	SD	0.0216	0.0141	0.0080	0.0103	0.0076	0.0027
	SDO	0.0594	0.0510	0.0315	0.0292	0.0176	0.0081
	$\overline{\mathrm{ASE}}$	0.0126	0.0176	0.0215	0.0271	0.0429	0.0583
$\hat{h}_{\mathrm{IPL}}$	Mean	0.1309	0.1374	0.1052	0.1139	0.0802	0.0493
	SD	0.0097	0.0101	0.0052	0.0076	0.0075	0.0033
	SDO	0.1137	0.0673	0.0451	0.0280	0.0181	0.0087
	$\overline{\mathrm{ASE}}$	0.0173	0.0196	0.0239	0.0268	0.0431	0.0588

Table 4: h_{M}, $\mathrm{M}(h_{\mathrm{M}})$ and statistics from 400 simulation ($n = 400$).

		m_0	m_1	m_2	m_3	m_4	m_5
h_{M}	True	0.2150	0.1625	0.1275	0.1000	0.0625	0.0425
	$\mathrm{M}(h_{\mathrm{M}})$	0.0065	0.0092	0.0116	0.0149	0.0237	0.0342
	$\overline{\mathrm{ASE}}$	0.0066	0.0085	0.0116	0.0143	0.0234	0.0336
h_{ASE}	Mean	0.2142	0.1642	0.1260	0.1012	0.0623	0.0427
	SD	0.0409	0.0338	0.0252	0.0186	0.0088	0.0053
	$\overline{\mathrm{ASE}}$	0.0060	0.0077	0.0105	0.0133	0.0226	0.0327
$\hat{h}_{\mathrm{RC}}$	Mean	0.2027	0.1506	0.1196	0.0934	0.0611	0.0418
	SD	0.0653	0.0455	0.0355	0.0238	0.0142	0.0087
	SDO	0.0834	0.0646	0.0477	0.0344	0.0196	0.0112
	$\overline{\mathrm{ASE}}$	0.0100	0.0125	0.0158	0.0179	0.0274	0.0381
$\hat{h}_{\mathrm{ODS}}$	Mean	0.1818	0.1530	0.1123	0.0954	0.0641	0.0421
	SD	0.0350	0.0289	0.0133	0.0135	0.0068	0.0030
	SDO	0.0665	0.0518	0.0335	0.0264	0.0139	0.0071
	$\overline{\mathrm{ASE}}$	0.0084	0.0108	0.0128	0.0159	0.0248	0.0346
$\hat{h}_{\mathrm{NDS}}$	Mean	0.1933	0.1635	0.1147	0.0993	0.0663	0.0425
	SD	0.0163	0.0104	0.0051	0.0054	0.0046	0.0015
	SDO	0.0529	0.0391	0.0298	0.0216	0.0130	0.0061
	$\overline{\mathrm{ASE}}$	0.0072	0.0089	0.0121	0.0147	0.0243	0.0339
$\hat{h}_{\mathrm{MDS}}$	Mean	0.1964	0.1655	0.1154	0.0998	0.0669	0.0431
	SD	0.0166	0.0110	0.0052	0.0054	0.0045	0.0015
	SDO	0.0521	0.0394	0.0295	0.0216	0.0131	0.0062
	$\overline{\mathrm{ASE}}$	0.0072	0.0089	0.0121	0.0147	0.0244	0.0339
$\hat{h}_{\mathrm{IPL}}$	Mean	0.1147	0.1204	0.0941	0.1008	0.0698	0.0437
	SD	0.0059	0.0055	0.0032	0.0040	0.0046	0.0018
	SDO	0.1084	0.0571	0.0412	0.0210	0.0146	0.0065
	$\overline{\mathrm{ASE}}$	0.0106	0.0101	0.0133	0.0146	0.0247	0.0341

Table 5. Empirical efficiencies (%) of h_{M} and of all bandwidth selectors.

	$n=50$							$n=100$						
	m_0	m_1	m_2	m_3	m_4	m_5	Mean	m_0	m_1	m_2	m_3	m_4	m_5	Mean
h_{M}	84	89	92	91	96	98	92	86	91	92	91	96	97	92
$\hat{h}_{\mathrm{RC}}$	49	58	62	67	76	88	67	48	64	62	69	75	84	67
$\hat{h}_{\mathrm{ODS}}$	56	68	71	73	82	92	74	55	74	75	80	83	92	77
$\hat{h}_{\mathrm{NDS}}$	63	76	76	79	83	93	78	69	85	81	84	86	95	83
$\hat{h}_{\mathrm{MDS}}$	66	78	79	80	82	93	80	71	85	82	84	86	95	84
$\hat{h}_{\mathrm{IPL}}$	50	66	70	79	84	95	74	52	75	76	85	87	94	78
	$n=200$							$n=400$						
	m_0	m_1	m_2	m_3	m_4	m_5	Mean	m_0	m_1	m_2	m_3	m_4	m_5	Mean
h_{M}	90	90	92	92	97	97	93	90	91	91	93	97	97	93
$\hat{h}_{\mathrm{RC}}$	58	67	72	72	82	84	73	60	62	67	74	83	86	72
$\hat{h}_{\mathrm{ODS}}$	70	76	81	82	89	93	82	71	71	82	83	91	95	82
$\hat{h}_{\mathrm{NDS}}$	83	86	88	88	91	95	89	83	87	87	90	93	96	89
$\hat{h}_{\mathrm{MDS}}$	83	86	88	88	91	95	89	83	87	87	90	93	96	89
$\hat{h}_{\mathrm{IPL}}$	60	77	79	89	91	95	82	56	77	79	91	92	96	82

Table 6. Empirical correlation coefficients of each bandwidth selector and h_{ASE}.

	$n=50$						$n=100$					
	m_0	m_1	m_2	m_3	m_4	m_5	m_0	m_1	m_2	m_3	m_4	m_5
$\hat{h}_{\mathrm{RC}}$	-.33	-.30	-.21	-.31	-.29	-.34	-.26	-.25	-.28	-.31	-.40	-.20
$\hat{h}_{\mathrm{ODS}}$	-.27	-.26	-.25	-.29	-.33	-.30	-.22	-.20	-.28	-.37	-.50	-.27
$\hat{h}_{\mathrm{NDS}}$	-.34	-.31	-.28	-.35	-.39	-.29	-.27	-.28	-.32	-.46	-.55	-.39
$\hat{h}_{\mathrm{MDS}}$	-.34	-.28	-.33	-.35	-.41	-.24	-.29	-.27	-.32	-.45	-.55	-.36
$\hat{h}_{\mathrm{IPL}}$	-.19	-.30	-.20	-.37	-.35	-.25	-.16	-.29	-.26	-.49	-.55	-.35
	$n=200$						$n=400$					
	m_0	m_1	m_2	m_3	m_4	m_5	m_0	m_1	m_2	m_3	m_4	m_5
$\hat{h}_{\mathrm{RC}}$	-.17	-.34	-.24	-.26	-.37	-.24	-.17	-.25	-.19	-.24	-.41	-.23
$\hat{h}_{\mathrm{ODS}}$	-.13	-.33	-.22	-.33	-.47	-.37	-.16	-.29	-.18	-.27	-.57	-.40
$\hat{h}_{\mathrm{NDS}}$	-.31	-.52	-.42	-.42	-.59	-.47	-.32	-.39	-.37	-.45	-.67	-.47
$\hat{h}_{\mathrm{MDS}}$	-.32	-.50	-.45	-.43	-.58	-.44	-.32	-.39	-.37	-.45	-.67	-.49
$\hat{h}_{\mathrm{IPL}}$	-.19	-.49	-.34	-.52	-.61	-.54	-.26	-.45	-.22	-.53	-.71	-.51

Appendix: Proofs of theorems

Proof of Proposition 2. In the following only a sketched proof for the use of $\hat{M}_{\mathrm{M}}$ will be carried out. This proof follows Härdle et al. (1992) and Feng (1999). Some details will be omitted to save space.

i). Observe that $\hat{b}(x_i)$ in (2.4) can be written as a linear combination of the observations

$$\hat{b}(x_i) = \sum_{j=1}^{n} A_j(x_i) Y_j \tag{A.1}$$

with the notation

$$A_j(x) = \sum_{k=1}^{n} w_{kh}(x) w_{jg}(x_k) - w_{jg}(x).$$

$\hat{M}_{\mathrm{M}}$ may be now decomposed into several components

$$\hat{M}_{\mathrm{M}}(h,g) = \hat{V}_{\mathrm{A}}(h) + B(h) + T_1 + T_2 + T_3 + 2T_4 + T_5, \tag{A.2}$$

where

$$\begin{aligned}
T_1 &= n^{-1} \sum_{i}^{*} \{[\sum_{j=1}^{n} m(x_j) A_j(x_i)]^2 - b(x_i)^2\}, \\
T_2 &= n^{-1} \sum_{i}^{*} \sum_{j=1}^{n} (\epsilon_j^2 - \sigma^2) A_j(x_i)^2, \\
T_3 &= n^{-1} \sum_{i}^{*} \sum_{j \neq k} \sum \epsilon_j \epsilon_k A_j(x_i) A_k(x_i), \\
T_4 &= n^{-1} \sum_{i}^{*} \sum_{j=1}^{n} \sum_{k=1}^{n} \epsilon_j m(x_k) A_j(x_i) A_k(x_i) \qquad \textit{and} \\
T_5 &= n^{-1} \sigma^2 \sum_{i}^{*} \sum_{j=1}^{n} A_j(x_i)^2.
\end{aligned}$$

T_2, T_3 and T_4 are all random variables with zero means. Consider bandwidths h such that $h/g \to 0$ as $n \to \infty$. Under this condition we obtain, following H"ardle et al. (1992) and Feng (1999),

$$\begin{aligned}
A_j(x_i) &= \sum_{k=1}^{n} w_{kh}(x_i)[w_{jg}(x_k) - w_{jg}(x_i)] \\
&\doteq n^{-2}(hg)^{-1} \sum_{k=1}^{n} K[(x_i - x_k)/h] f^{-1}(x_k) \{L[(x_k - x_j)/g] \\
&- L[(x_i - x_j)/g]\} f^{-1}(x_j) \\
&\doteq (ng)^{-1} f^{-1}(x_j) \int K(y) \{L[(x_i - x_j)/g - hg^{-1}y] - L[(x_i - x_j)/g]\} dy \\
&\doteq \kappa_r (ng)^{-1} (h/g)^r f^{-1}(x_j) L^{(r)}[g^{-1}(x_i - x_j)] \\
&\doteq \kappa_r (ng)^{-1} (h/g)^r f^{-1}(x_i) L^{(r)}[g^{-1}(x_i - x_j)].
\end{aligned}$$

Based on this approximation it can be shown that

$$T_1 = O(h^{2r} g^s),$$

$$T_5 = O(\frac{1}{nh})(\frac{h}{g})^{(2r+1)}.$$

Furthermore it can be shown that T_4 is asymptotically negligible and

$$T_2 = o_p(T_3) = o_p(T_5).$$

Note that $\hat{V}_{\mathrm{A}} = O_p(n^{-1}h^{-1})$ and that, in this case, $B(h) = o_p(\hat{V}_{\mathrm{A}}(h))$ and $T_1 = o(T_5)$, we have finally

$$\hat{M}_{\mathrm{M}}(h, g) \doteq O_p(n^{-1}h^{-1}) + O_p(\frac{1}{nh})(\frac{h}{g})^{(2r+1)}. \tag{A.3}$$

This means that, for $h = o(g)$, $\hat{M}_{\mathrm{M}}(h)$ decreases monotonously in h in probability and so that the minimizer of $\hat{M}_{\mathrm{M}}$ is at least of order $O(g)$. i) is proved.

ii). Note that now $g = O_p(h_{\mathrm{M}}) = O_p(n^{-1/(2r+1)})$. At first, it can be shown, similarly to the analysis for case i), that, for $h = o(h_{\mathrm{M}})$, $\hat{M}_{\mathrm{M}}(h)$ decreases monotonously in h in probability. For a bandwidth $h = O(h_{\mathrm{M}})$ it is easily to show that, in this case, $\hat{M}_{\mathrm{M}}(h) = O_p(n^{-2r/(2r+1)})$. Furthermore, if h is a bandwidth such that $h_{\mathrm{M}} = o(h)$, then $\hat{M}_{\mathrm{M}}$ will be dominated by the bias term $B(h) = O(h^{2r})$, which is of larger order than $O_p(n^{-2r/(2r+1)})$ and increases monotonously in h. This implies $\hat{h} = O_p(h_{\mathrm{M}})$.

iii). In this case Theorem 1 in Heiler and Feng (1998) holds due to the assumption $h_{\mathrm{M}} = o_p(g)$. The result hence holds following that theorem. The exact rate of convergence depends however on the order of g.

Proof of Theorem 1: Note that the IDS procedure is just a special DS method with a fixed pilot bandwidth $g = n^{\alpha} h_{\mathrm{M}}$, when convergence is reached. Hence we obtain the results of Theorem 1 by inserting this fixed pilot bandwidth into Theorem 1 in Heiler and Feng (1998).

Proof of Theorem 2:

1. The proof of 1 is straightforward and is omitted.

2. Now we just need to calculate the variance σ_2^2. The proof follows Härdle et al. (1988, 1992), Herrmann (1994) and Feng (1999).

   Following Härdle et al. (1988) we have

$$\hat{h}_{\mathrm{MDS}} = h_{\mathrm{M}} - (\hat{M}'_{\mathrm{M}}(h_{\mathrm{M}}) - M(h_{\mathrm{M}}))/M''(h^*), \tag{A.4}$$

   where h^* is between h_{ASE} and h_{M}. Using (A.2) it can be shown that the stochastic part of $\hat{M}'_{\mathrm{M}}(h_{\mathrm{M}}) - M(h_{\mathrm{M}})$ is dominated by $\hat{V}'_{\mathrm{A}}$, T'_3 and $2T'_4$. For $\hat{h}_{\mathrm{IPL}}$, Herrmann (1994) showed that the stochastic part of $\hat{M}'_{\mathrm{A}}(h_{\mathrm{M}}) - M(h_{\mathrm{M}})$ is dominated by $\hat{V}'_{\mathrm{A}}$ and another term, say $2T'_6$, which, in the current context, has the form

$$T'_6 = \frac{1}{n} h_{\mathrm{A}} I(K)^2 \sum_{j=1}^{n} m^{(4)}(x_j) \mathbb{1}_{[c,d]}(x_j)\epsilon_j. \tag{A.5}$$

   Following Härdle et al. (1992) and Feng (1999), we have

$$T'_4 \doteq \frac{1}{n} h_{\mathrm{A}} I(K)^2 \beta \sum_{j=1}^{n} m^{(4)}(x_j) \mathbb{1}_{[c,d]}(x_j)\epsilon_j. \tag{A.6}$$

   This means that $T'_4 = T'_6(1 + o(1))$. Furthermore, it can be shown that both, $\hat{V}'_{\mathrm{A}}$ and T'_4, are asymptotically independent of T'_3. Hence we have

$$\mathrm{Var}[n^{7/10}(\hat{h}_{\mathrm{MDS}} - h_{\mathrm{M}} - O(n^{-29/45}))] = \mathrm{Var}[n^{7/10}(\hat{h}_{\mathrm{IPL}} - h_{\mathrm{M}} - O(n^{-2/5}))] + \gamma_2. \tag{A.7}$$

   Using the results on $\hat{h}_{\mathrm{IPL}}$ in Herrmann (1994) we obtain the formula for σ_2^2.

3. Using (A.13) in Härdle et al. (1988), it can be shown that $h_{\mathrm{ASE}} - h_{\mathrm{M}}$ is asymptotically independent of T'_3 too, due to the fact that $h_{\mathrm{M}}/g \to \infty$. This means that the covariance between $\hat{h}_{\mathrm{MDS}}$ and h_{ASE} is the same as the one between $\hat{h}_{\mathrm{IPL}}$ and h_{ASE} given in Herrmann (1994). Hence Theorem 2 is proved.