DFG-SNF Research Group FOR916
Statistical Regularization and Qualitative Constraints

Enno Mammen
Maria Dolores Martinez Miranda
Jens Perch Nielsen
Stefan Sperlich

Do-validation for Kernel Density Estimation

Preprint FOR916 10-22

Preprint-Series of the Research Group FOR916

FOR916/10-22 March 2010

Do-validation for Kernel Density Estimation

Enno Mammen

Universität Mannheim, Abteilung Volkswirtschaftslehre

L7, 3-5, 68131-Mannheim

María Dolores Martínez Miranda

Universidad de Granada, Departamento de Estadística e I.O.

Campus Fuentenueva, E - 18071 Granada

Jens Perch Nielsen

Cass Business School, City University

106 Bunhill Row, UK - London EC1Y 8TZ

Stefan Sperlich

Georg-August Universität Göttingen, Institut für Statistik und Ökonometrie

Platz der Göttinger Sieben 5, D - 37073 Göttingen

March 8, 2010

Abstract

We consider the fundamental model selection problem of mathematical statistics. The study of bandwidth selection in kernel density estimations took major steps forward with the papers of Hall and Marron (1987), Hall and Johnstone (1992). Since then focus seems to have been on various versions of implementing the so called plug-in method aiming at estimating the minimum integrated squared error. The most successful of these efforts still seem to be the plug-in method of Sheather and Jones (1991) that we also use as overall benchmark in this paper. In this paper we derive a new theorem deriving the asymptotic theory of combinations of directly crossvalidated bandwidths, indirectly crossvalidated bandwidths and plug-in bandwidths , where we take advantage of recent advances in the study of indirect crossvalidation, see Hart and Yi (1998), Hart and Li (2005) and Savchuk, Hart and Sheather (2010a,b). We conclude that the slow convergence of asymptotic theory for bandwidth selectors implies that once asymptotic theory is close to that of plug-in then it is the practical implementation that counts. This insight led us to a bandwidth selector search with the symmetrized density version of onesided crossvalidation as a clear winner.[1]

Keywords: bandwidth choice, crossvalidation, plug-in, nonparametric estimation.

---

[1] This research was financially supported by knowledge company Festina Lente, Deutsche Forschungsgemeinschaft (DFG), Projects FOR916, MA1026/7-2 and MA1026/9-1 and MTM2008-03010/MTM.

1 Introduction

Standard (least-squares) crossvalidation was proposed by Rudemo (1982) and Bowman (1984). The simplicity of its implementation and its intuitively appealing interpretation probably makes it the most popular automatic bandwidth selection method. Its practical data driven flavor makes up for its lack of stability in the eyes of many practitioners. The lack of stability of standard crossvalidation has been pointed out for both kernel density estimation and kernel regression estimation (Hall and Marron, 1987; Härdle, Hall and Marron, 1988). These drawbacks of standard crossvalidation have motivated several studies on more stable bandwidth selectors, most of them related to the plug-in method (Biased crossvalidation by Scott and Terrell, 1987; smoothed crossvalidation by Hall, Marron and Park, 1989; Sheather and Jones, 1991; the stabilized bandwidth selector rule by Chiu, 1991; the kernel contrast method of Ahmad and Ran, 2004; recent bootstrap methods, see Chacón, Montanero, and Nogales, 2008; among others). Better performance of plug-in has been questioned in Loader (1999). He points out that the value of the chosen bandwidth heavily depends on the arbitrary specification of pilot bandwidths and may be strongly biased in case of misspecification. Still quite recently, SiZer became a popular alternative, see Chaudhuri and Marron (1999), Godtliebsen, Marron and Chaudhuri (2002), and Hanning and Marron (2006). Note that SiZer does not search for an optimal data driven bandwidth, it rather highlights for each bandwidth which features (of the density or regression) get detected. In Żychaluk and Patil (2008) it has been argued that instability of cross validation is often due to discretization effects. Another method that improves on ordinary crossvalidation is onesided crossvalidation. Onesided crossvalidation is one version of the general principle of indirect crossvalidation, see Savchuk, Hart and Sheather (2010a). Onesided crossvalidation was originally proposed by Hart and Yi (1998) for local linear regression and was extended to our density case by Martínez-Miranda, Nielsen and Sperlich (2009). In this paper we derive a new theorem on combinations of kernel density bandwidths and investigate its practical potential. This new theorem allows combinations of bandwidths from both direct and indirect crossvalidation and can include the plug-in bandwidth as well. While we have investigated hundreds of combinations through simulation studies, we only present those eight bandwidth combinations that are sufficient for our overall conclusions. First we look at the optimal combinations of the standard crossvalidation window and the plug-in window. With the Epanechnikov kernel it turns out that the optimal combinations pick 1.21 times the plug-in window and place a negative weight of 0.21 of the crossvalidation bandwidth. This is not surprising considering the well known negative correlation between the crossvalidation window and the plug-in window, see Hall and Marron (1987). We compare the unfeasible optimal plug-in bandwidth with this new and better combined bandwidth and the new combined bandwidth is slightly better. However, when considering the same question for feasible bandwidths, this new asymptotically superior bandwidth does not impress in finite sample studies. The reason seems to be that feasible plug-in estimators have a tendency to oversmooth in finite samples while standard crossvalidation is almost unbiased. Our new optimal combination of plug-in and crossvalidation suffers in practice from a tendency to oversmooth even more. From this study we learn that the slow convergence of bandwidth selection asymptotics does have that consequence that we can not fully rely on theory when picking our practical bandwidth selector. To illustrate this point even more, we implemented the intuitive simple average of the plug-in bandwidth and the crossvalidation bandwidth. This simple combinations is appealing because of the well known practical experience of feasible plug-in bandwidths to oversmooth, while the almost unbiased crossvalidation bandwidth sometimes have very bad performance because of undersmoothing. Therefore a simple average should intuitively improve both. And this turns indeed out to be the case. The asymptotic performance of this simple average is much better than for ordinary crossvalidation and only slightly worse than for the plug-in method. In practice it beats both. This insight led to a search for good finite sample performance of combinations of bandwidths with an asymptotic performance close to the plug-in method. The overall winner of this search was a simple combination of the right-sided and the left-sided versions of onesided crossvalidation. We call this new bandwidth selector "do-validation" (double onesided crossvalidation). The conclusion of this paper therefore suggests the do-validation bandwidth as an asymptotically well performing bandwidth selector with excellent finite sample properties.

Let $\widehat{f}_{h,K}$ be some density estimator based on the random sample $X_1, X_2, \ldots, X_n$ from a distribution with density function, $f(\cdot)$, h is the bandwidth parameter and K the kernel.

Then the integrated squared error (ISE) can be written as

$$\Delta_K(h) = \int \left\{ \widehat{f}_{h,K}(x) - f(x) \right\}^2 dx$$

The crossvalidation approach estimates $\Delta_K(h)$ by

$$\widehat{\Delta}_K(h) = \int \left[\left\{ \widehat{f}_{h,K}(x) \right\}^2 + f(x)^2 \right] dx \; - 2 \int \widehat{f}^{(i)}_{h,K}(x) d\widehat{F}(x),$$

where $\widehat{f}^{(i)}_{h,K}$ is the version of $\widehat{f}_{h,K}$ where the i'th observation is left out in the estimation such that

$$\int \widehat{f}^{(i)}_{h,K}(x) d\widehat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} \widehat{f}^{(i)}_{h,K}(X_i), \tag{1}$$

and $\widehat{F}$ is the empirical cdf estimating the cdf F corresponding to the density f. Onesided crossvalidation does not work in the local constant case because of the inferior rate of convergence of the onesided kernel density estimator based on a local constant kernel. Therefore onesided crossvalidation has to based on local linear density estimation as suggested by Martínez-Miranda, Nielsen and Sperlich (2009). It turns out that left-onesided crossvalidation and right-onesided crossvalidation are not identical in the local linear case (as in in the local constant case), this is because the local linear corrections differ. However, asymptotically local linear left-onesided crossvalidation and local linear right-onesided crossvalidation are equivalent. We therefore get the same asymptotic theory for their simple average (do-validation) as for each of them separately. Left-onesided and right-onesided crossvalidation also behave more or less in the same way for most densities. For asymmetric densities this is, however, not the case. When we for example have a long tailed density then the asymmetric bias correction of the left-onesided compared to the right-onesided crossvalidation leads to considerable different performance in practice. Do-validation delivers a good stable compromise. It has the same asymptotic theory as each of the two onesided alternatives and a better overall finite sample performance.

In Section 2 we define the new class of combined crossvalidation bandwidth selectors. In Section 3 the theoretical properties of this class of bandwidths is derived and the opportunity of adding the plug-in bandwidth to the combinations is included. We consider eight bandwidth selectors in Section 4: standard crossvalidation, one version of onesided crossvalidation, do-validation, the unfeasible optimal plug-in method and its feasible implementation, the unfeasible optimal combination of crossvalidation and plug-in and its feasible implementation, and finally the simple average of the crossvalidation bandwidth and the feasible plug-in bandwidth. For all considered bandwidths there are two components of noise in the asymptotic performance. While standard crossvalidation and the plug-in method have the first noise component in common, it turns out that the second noise component is very different and in favor of the plug-in method. All our eight considered bandwidths have the same first noise component as classical crossvalidation and the plug-in method have. The second noise component varies with the different methods. While it is quite big for the crossvalidation method, then when cut to one third - as our practical alternatives do - then its exact value becomes less important than the practical performance of the implementation at hand. In Section 5 we go through a finite sample study that shows that do-validation is the best of considered method for finite samples and is slightly better that its simple onesided alternatives, the simple average of the plug-in method and crossvalidation comes third. The feasible plug-in method only comes fourth while still performing much better than classical crossvalidation that looses the test.

2 A general class of data-driven bandwidth estimators

Martínez-Miranda, Nielsen and Sperlich (2009) proposed the onesided crossvalidation method (OSCV hereafter) for kernel density estimation. The authors considered the local linear kernel density estimator (Jones, 1993; and Cheng 1997a, 1997b), which can be defined by the equivalent-kernel expression

$$\widehat{f}_{h,K}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \tag{2}$$

where $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with density function, $f(\cdot)$, h is the bandwidth parameter and the equivalent kernel is given by

$$K(u) = \frac{k_2 - k_1 u}{k_0 k_2 - (k_1)^2} K_0(u), \tag{3}$$

with K_0 being a kernel function with finite moments $k_l = \int u^l K_0(u) du$ $(l = 0, 1, 2)$.

Let us consider any common symmetric kernel function, K_0, satisfying $k_0 = 1$, $k_1 = 0$ and $k_2 < \infty$, and its left-onesided version

$$K_1(u) = \begin{cases} 2K_0(u) & \text{if} \quad u < 0, \\ 0 & \text{otherwise.} \end{cases}$$

For the kernel K_1 the local linear kernel correction is

$$L_1(u) = \frac{k_{2,1} - u k_{1,1}}{k_{0,1} k_{2,1} - k_{1,1}^2} K_1(u),$$

where $k_{l,1} = \int t^l K_1(t) dt$ $(l = 0, 1, 2)$. Simple calculations yield $k_{0,1} = 1$, $k_{2,1} = k_2$ and, therefore we have

$$L_1(u) = \frac{k_2 - u k_{1,1}}{k_2 - (k_{1,1})^2} 2K_0(u) \mathbf{1}_{(-\infty,0)}. \tag{4}$$

Now the left-onesided local linear kernel density estimator, $\widehat{f}_{h,L_1}$, can be defined as in (2) but with K replaced by L_1.

With these definitions the left-OSCV criterion (OSCV_1) is defined by

$$\mathrm{OSCV}_1(h) = \int \widehat{f}^{\,2}_{h,L_1}(x)dx - 2n^{-1}\sum_{i=1}^{n} \widehat{f}_{h,L_1}(X_i), \tag{5}$$

with $\widehat{h}_1$ as its minimizer; and the left-OSCV bandwidth is calculated from $\widehat{h}_1$ by

$$\widehat{h}_{1,\mathrm{OSCV}} = \mathrm{C}\widehat{h}_1,$$

where

$$\mathrm{C} = \left(\frac{c_1}{c_2}\frac{c_{2,1}}{c_{1,1}}\right)^{1/5}, \tag{6}$$

with $c_1 = \int K^2(x)dx,\ c_2 = \left(\int x^2K(x)dx\right)^2 \equiv k_2^2,\ c_{1,1} = \int L_1^2(x)dx$ and $c_{2,1} = \left(\int x^2L_1(x)dx\right)^2$, see Martínez-Miranda, Nielsen and Sperlich (2009).

Exactly the same way we define the right-onesided version of K_0 by

$$K_2(u) = \begin{cases} 2K_0(u) & \text{if } \ u > 0 \\ 0 & \text{otherwise} \end{cases}$$

the right-onesided local linear kernel by

$$L_2(u) = \frac{k_2 - 2u\int_0^\infty tK_0(t)dt}{k_2 - \left(2\int_0^\infty tK_0(t)dt\right)^2}2K_0(u)\mathbf{1}_{(0,\infty)} = \frac{k_2 + uk_{1,1}}{k_2 - (k_{1,1})^2}2K_0(u)\mathbf{1}_{(0,\infty)}, \tag{7}$$

and the right-onesided local linear density estimator, $\widehat{f}_{h,L_2}$ based on L_2. Then the right-OSCV criterion, OSCV_2, is defined as (5) but with $\widehat{f}_{h,L_1}$ replaced by $\widehat{f}_{h,L_2}$. The right-OSCV bandwidth is calculated by $\widehat{h}_{2,\mathrm{OSCV}} = \mathrm{C}\widehat{h}_2$, where C is the same as in (6) and $\widehat{h}_2$ is the minimizer of OSCV_2.

The bandwidth selectors in the class are based on the inspection of kernel density estimators $\widehat{f}_{h,L_j}(x), 1 \leq j \leq J$, for kernels L_j that fulfill $L_j(0) = 0$. For $1 \leq j \leq J$ we define bandwidth selectors $\widehat{h}_j$ by the crossvalidation method

$$\widehat{h}_j = \arg\min_h \int \widehat{f}_{h,L_j}(x)^2 dx - 2n^{-1} \sum_{i=1}^n \widehat{f}_{h,L_j}(X_i).$$

Note that because of $L_j(0) = 0$ we have not to use a leave-one-out version of $\widehat{f}_{h,L_j}$ in the sum on the right hand side. For some weights w_j (not necessary being positive) with $\sum_{j=1}^J w_j = 1$, a new bandwidth selector $\widehat{h}$ is defined by

$$\widehat{h} = \sum_{j=1}^J w_j \left(\frac{c_1}{c_2} \frac{c_{2,j}}{c_{1,j}} \right)^{1/5} \widehat{h}_j \tag{8}$$

with $c_1 = \int K^2(x)dx$, $c_2 = \left(\int x^2 K(x)dx\right)^2$, $c_{1,j} = \int L_j^2(x)dx$, $c_{2,j} = \left(\int x^2 L_j(x)dx\right)^2$ $(1 \leq j \leq J)$.

Note that the do-validation comes from this general class putting $J = 2$, $w_j = 1/2$, $j = 1, 2$, choosing K as the local linear kernel given in (3) from a symmetric kernel K_0, and L_1, L_2, as the left- and right-onesided local linear kernels given in (4) and (7), respectively.

3 Asymptotic theory

Here we compare $\widehat{h}$ with the MISE-optimal bandwidth h_0 with the ISE-optimal bandwidth $\widetilde{h}$,

$$\begin{aligned} h_0 &= \arg\min_h \mathrm{E}\left[\int \left(\widehat{f}_{h,K}(x) - f(x)\right)^2 dx\right], \\ \widetilde{h} &= \arg\min_h \left[\int \left(\widehat{f}_{h,K}(x) - f(x)\right)^2 dx\right]. \end{aligned}$$

Under our assumptions, see below, it holds that $h_0 = \left(\frac{c_1}{c_2 R_{f''}}\right)^{1/5} n^{-1/5} + o\left(n^{-3/10}\right)$, with $R_{f''} = \int f''^2(x)dx$. In general we define a generic $R_g = \int g^2(x)dx$ for square-integrable functions g.

We now state a theorem about the asymptotic distribution of $\widehat{h} - h_0$ and $\widehat{h} - \widetilde{h}$. We are also interested in $\widehat{h}_0 - h_0$, with $\widehat{h}_0$ being an unfeasible combination with the oracle bandwidth h_0 defined by

$$\widehat{h}_0 = \sum_{j=1}^{J-1} w_j \left(\frac{c_1}{c_2}\frac{c_{2,j}}{c_{1,j}}\right)^{1/5} \widehat{h}_j + w_J h_0 \tag{9}$$

where $\sum_{j=1}^{J} w_j = 1$). For this result we need the following assumptions:

(A1) K and L_j ($j = 1, \ldots, J$) are compactly supported. The kernels are differentiable on $I\!R - \{0\}$ and have derivatives that are Hölder continuous on $I\!R - \{0\}$ i.e. there exist constants $c, \delta > 0$ such that $|g(x) - g(y)| \leq c|x - y|^\delta$ for $x, y < 0$ or $x, y > 0$ with g equal to K' or L_j' ($j = 1, \ldots, J$). For $j = 1, \ldots, J$ it holds that $L_j(0) = 0$ and that $\int uL_j(u)du = 0$ and $\int uK(u)du = 0$.

(A2) The density f is bounded and twice differentiable. The derivatives f' and f'' are bounded and integrable. The second derivative is Hölder continuous with exponent $\delta > \frac{1}{2}$.

For do-validation condition (A1) is satisfied if the kernel K_0 satisfies the required smoothness conditions.

Theorem 1. ***Combination of bandwidths.*** *Under A1, A2 the bandwidth selector*

$\widehat{h}$ in (8) satisfies that

$$\begin{aligned}
n^{3/10}(\widehat{h}-\widetilde{h}) &\to N(0,\sigma_1^2) \quad \textit{in distribution,} \\
n^{3/10}(\widehat{h}-h_0) &\to N(0,\sigma_2^2) \quad \textit{in distribution,}
\end{aligned}$$

where

$$\begin{aligned}
\sigma_k^2 = \;& \frac{4}{25} c_1^{-2/5} c_2^{-3/5} R_{f''}^{-8/5}\,\mathrm{V}(f'')\delta_k \\
& + \frac{1}{50} c_1^{-7/5} c_2^{-3/5} R_{f''}^{-3/5} R_f \int \left[\delta_k H(u) - \sum_{j=1}^{J} w_j \left(\frac{c_1}{c_{1,j}}\right) H_j(u)\right]^2 du,
\end{aligned} \tag{10}$$

with

$$\mathrm{V}(f'') = \int f''^2(x) f(x)dx - \left(\int f''(x)f(x)dx\right)^2,$$

$$\begin{aligned}
H(u) &= 2\int K(u+v)K(v)dv + 2\int K(-u+v)K(v)dv \\
&\quad +2\int K(u+v)vK'(v)dv + 2\int K(-u+v)vK'(v)dv,
\end{aligned}$$

$$\begin{aligned}
H_j(d_j u) &= 2\int L_j(u+v)L_j(v)dv + 2\int L_j(-u+v)L_j(v)dv \\
&\quad +2\int L_j(u+v)vL_j'(v)dv + 2\int L_j(-u+v)vL_j'(v)dv \\
&\quad -2\left[L_j(u) + uL_j'(u) + L_j(-u) - uL_j'(-u)\right],
\end{aligned}$$

$$d_j = \left(\frac{c_1}{c_{1,j}}\frac{c_{2,j}}{c_2}\right)^{-1/5}, \qquad \delta_k = \begin{cases} 1 & \textit{for } k=1, \\ 0 & \textit{for } k=2. \end{cases}$$

Also the unfeasible combination $\widehat{h}_0$ in (9) satisfies that

$$n^{3/10}(\widehat{h}_0-\widetilde{h}) \to N(0,\sigma_1^2) \quad \textit{in distribution,}$$

with σ_1^2 as in 10 but defining the last $H_J = 0$.

Under additional smoothness assumption on f, Hall and Johnstone (1992) discussed efficient estimation of the ISE-optimal bandwidth $\widetilde{h}$. They showed that estimation of $\widetilde{h}$ is asymptotically equivalent to the estimation of $R_{f'} = \int f'(x)^2 dx$. Using an efficient estimator of $R_{f'}$ one gets an estimator $\widehat{h}$ of $\widetilde{h}$ where $n^{3/10}(\widehat{h} - \widetilde{h})$ has asymptotic variance $\frac{4}{25} c_1^{-2/5} c_2^{-3/5} R_{f''}^{-8/5} \mathrm{V}(f'')$. Thus, in our class of bandwidth selectors a bandwidth would achieve the optimality bound if

$$\int \left[H(u) - \sum_{j=1}^{J} w_j \left(\frac{c_1}{c_{1,j}} \right) H_j(u) \right]^2 du \quad = \quad 0. \tag{11}$$

We do not know if this claim can be achieved by appropriate choice of kernels L_j and weights w_j, this is our purpose in next section.

4 Eight competing combination of bandwidth selectors

For given kernels K and L_j $(j = 1, \ldots, J)$ and a density f, we might explicitly calculate the components in the asymptotic variance in (10), for the just introduced class of bandwidth selectors (8), and also for the unfeasible combinations (9). We look for good selectors, in the optimality sense of Hall and Johnstone (1992) i.e. bandwidths which, holding the first variance term in (10), achieve that the second term gets small. Hall and Johnstone (1992) showed that an asymptotically achievable bandwidth exists and that the definition of such an optimal bandwidth is that the asymptotic second noise component is zero. While Hall and Johnstone (1992) indicated that such optimal methods do indeed seem to exist, they never pursued the issue any further and did not provide practical examples of such bandwidth selectors. In our search for good bandwidth selectors, we do for the first time provide a feasible bandwidth selector with better asymptotic theory than the plug-in method, namely the optimal combination of the feasible plug-in method (with weight w_1) and classical crossvalidation. Figure 1 shows the graphs of the resultant second variance term against the weight. We plot the noisy term in a wide range including negative weights to get optimality (as we argued in Section 1 because of the known negative correlation between them). And indeed the optimum is achieved by weighting the plug-in bandwidth with $w_1 = 1.2116$ and crossvalidation with $w_2 = 1 - w_1 = -0.2116$. Such optimal combination yields a second noisy term of 0.5059 (lower than 0.7169 which is provided by the oracle plug-in h_0 as we show below).

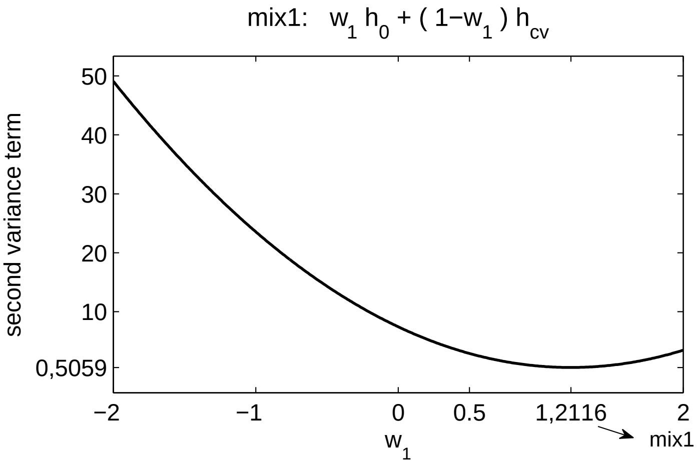

Figure 1: Second component of the asymptotic variance corresponding to combine plug-in bandwidth with standard crossvalidation. The optimal value is achieved by weighting the plug-in bandwidth with $w_1 = 1.2116$.

But we show in our finite sample section that when we are at this level of excellent asymptotic performance, then it becomes irrelevant to improve the asymptotic performance any further. Then it is the performance of practical implementation that counts.

We consider eight different bandwidths. Left-onesided crossvalidation and the simple average of left-sided and right-sided crossvalidation: do-validation method. The standard crossvalidation bandwidth $\widehat{h}_{\mathrm{CV}}$ and the MISE-optimal h_0 (the oracle plug-in method), also defined in Theorem 2.1 by Hall and Marron (1987). We also consider three more combinations: the optimal combination of the oracle plug-in and the standard CV bandwidth, given by $\widehat{h}_{\mathrm{mix1}} = 1.2116h_0 - 0.2116\widehat{h}_{\mathrm{CV}}$; the feasible version $\widehat{h}_{\mathrm{mix2}} = 1.2116h_{PI} - 0.2116\widehat{h}_{\mathrm{CV}}$, with h_{PI} the plug-in selector by Sheather and Jones (1991); and finally the pragmatic average $\widehat{h}_{\mathrm{mix3}} = 0.5h_{PI} + 0.5\widehat{h}_{\mathrm{CV}}$. Note that asymptotic properties of h_{PI} and the oracle h_0 are the same (and therefore those of mix1 and mix2).

Considering the Epanechnikov kernel, the asymptotic variances of $n^{3/10}(\widehat{h} - \widetilde{h})$ and $n^{3/10}(\widehat{h}_0 - \widetilde{h})$ in Theorem 1, expression (10), for each of the eight considered bandwidths are given by

$$
\begin{aligned}
\sigma^2_{\mathrm{OSCV}} &= C_{f,K}\left\{4c_1\frac{\mathrm{V}(f'')}{R_{f''}R_f}+2.1884\right\}\\
\sigma^2_{\mathrm{DO}} &= C_{f,K}\left\{4c_1\frac{\mathrm{V}(f'')}{R_{f''}R_f}+2.1884\right\}\\
\sigma^2_{\mathrm{CV}} &= C_{f,K}\left\{4c_1\frac{\mathrm{V}(f'')}{R_{f''}R_f}+7.4234\right\}\\
\sigma^2_{0} &= C_{f,K}\left\{4c_1\frac{\mathrm{V}(f'')}{R_{f''}R_f}+\ 0.7169\right\}\\
\sigma^2_{\mathrm{mix1,2}} &= C_{f,K}\left\{4c_1\frac{\mathrm{V}(f'')}{R_{f''}R_f}+\ 0.5059\right\}\\
\sigma^2_{\mathrm{mix3}} &= C_{f,K}\left\{4c_1\frac{\mathrm{V}(f'')}{R_{f''}R_f}+\ 2.8921\right\}
\end{aligned}
$$

with $C_{f,K}=\frac{1}{25}c_1^{-7/5}c_2^{-3/5}R_{f''}^{-3/5}R_f$.

Therefore there are two components which inflate all the variances, but only the second term differs from each other selector. We can observe a clear reduction in this second variance term by both onesided crossvalidation and do-validation, compared with standard crossvalidation. The asymptotic variance of the oracle plug-in method is lower our practical alternatives. However, it is beaten by the optimal combination of itself with classical crossvalidation. But as we will see in the next section, the asymptotic theory of all our considered bandwidths - except for classical crossvalidation - is now so good that it becomes irrelevant. Now it is the practical performance that matters.

5 Finite Sample Performance

While the above presented theoretical results show the asymptotic good performance of some combinations of CV bandwidths, this section is to study its performance in practice when we face finite samples, sometimes just of moderate size. Thereby we will compare the following bandwidths: standard crossvalidation (CV hereafter), left- and right- onesided crossvalidation (OSCV$_1$ and OSCV$_2$ hereafter), the do-validation method, the three combinations mix1, mix2 and mix3 defined in Section 4, and also the feasible type plug-in bandwidth (PI) by Sheather and Jones (1991) and the oracle plug-in.

This plug-in bandwidth PI is calculated from the asymptotic expression of the MISE-optimal bandwidth, $h_0 \approx \left(\frac{c_1}{c_2 R_{f''}}\right)^{1/5} n^{-1/5}$ where c_1 and c_2 are known whereas $R_{f''}$ has to be estimated with a prior bandwidth g_p. To this aim, take Silverman's rule of thumb bandwidth g_p for Gaussian kernels where the standard deviation of X is estimated by the minimum of two methods: the moment estimate s_n and the interquartile range IR_X divided by 1.34, i.e. $g_S = 1.06\ \min\{IR_X 1.34^{-1}, s_n\} n^{-1/5}$. As the quartic kernel K_Q comes close to the Epanechnikov but allows for estimating the second derivative, we normalize g_S by the factors of the canonical kernel (Gaussian to quartic) and adjust for the slower rate ($n^{-1/9}$) needed to estimate second derivatives, i.e.

$$g_p = g_S \frac{2.0362}{0.7764}\, n^{1/5-1/9} .$$

Next,

$$\widehat{R}_{f''} = \int \widehat{f}''^2 - \frac{1}{n g_p^5} \int K_Q''^2$$

to correct for the bias inherited by

$$\widehat{f}''(x) = \frac{1}{n g_p^3} \sum_{i=1}^{n} K_Q'' \left(\frac{X_i - x}{g_p} \right) .$$

In simulation studies not shown here this prior choice turned out to perform better than any of the many other feasible plug-in methods we tried. At least for the estimation problems discussed in this paper. There is another recent interesting alternative to standard plug-in implementations, namely the indirect crossvalidation alternative of Savchuk, Hart and Sheather (2010a). One could interpret their bandwidth selector as a feasible plug-in selector because it is asymptotically equivalent to the plug-in method in the theoretical sense considered in this paper. We could also call their method a pilot-free plug-in method, because they obtain this equivalence to the plug-in bandwidth without any pilot density to plug-in. The key idea of their paper is to do indirect crossvalidation having two components in the first indirect kernel, one with a stable bias and variance and another with an exploding variance and a bias going to zero. While pilot-free plug-in estimation does not need any pilot density, it does need to determine somehow a good trade off between the two entering components. We tried a number of versions of pilot-free plug-in estimation. But since we did not achieve serious practical gains compared to Sheather and Jones (1991), we decided to stick to the more traditional and well known plug-in estimator based on a pilot density estimator. The simulation study in Savchuk, Hart and Sheather (2010b) gives similar conclusion as ours for the indirect crossvalidation alternative of Savchuk, Hart and Sheather (2010a) : current implementations do not imply practical improvements to Sheather and Jones (1991). However, we consider pilot-free plug-in estimation an important area of future research in kernel density bandwidth selection. The key research element of pilot-free plug-in estimation seems to be to determine a practical solution to the trade off between the two entering components in the indirect kernel. The discussion on pilot-free plug-in estimation also give us some intuition to why onesided crossvalidation and do-validation work so well. While both these two latter methods are practical and easy to implement, they are a first step towards a pilot-free plug-in estimator. In the first step it does blow up variance to some extent while keeping bias relatively stable. However, it does not bother to optimize this idea asymptotically by blowing up the variance indefinitely in the the indirect step with all the practical problems this implies. One-sided crossvalidation and do-validation are practical and pragmatic first steps towards plug-in free pilot estimation, but with easy stable implementations that work extremely well in practice. See also Hart and Li (2005) for more intuitive insight to why onesided crossvalidation works, they consider the nonparametric regression case of Hart and Yi (1998).

When we mix for do-validation the OSCV_1 and OSCV_2 bandwidths, OSCV_1 and OSCV_2, are expected to give quite different results in practice when the density to be estimated is asymmetric and/or has boundary problems. For this reason we will carry out the comparison over a set of densities comprising one to three modes, strong asymmetries and some boundary problems (not abrupt but exponentially falling tails, otherwise one has to work with boundary correcting kernels anyway). More specific, our selected data generating processes (DGP) are the following six normal, gamma and mixed densities, see also Figure 2:

1. a simple normal distribution, $N(0.5, 0.2^2)$,
2. a mixture of two normals which were $N(0.35, 0.1^2)$ and $N(0.65, 0.1^2)$,
3. and a mixture of three normals, namely $N(0.25, 0.075^2)$, $N(0.5, 0.075^2)$ and $N(0.75, 0.075^2)$ giving three clear modes.

4. Further, a gamma, $Gamma(a, b)$ with $b = 1.5$, $a = b^2$ applied on $5x$ with $x \in I\!\!R_+$, i.e.

$$f(x) = 5\frac{b^a}{\Gamma(a)}(5x)^{a-1}e^{-5xb},$$

5. a mixture of two gamma, $Gamma(a_j, b_j)$, $a_j = b_j^2$, $b_1 = 1.5$, $b_2 = 3$ applied on $6x$, i.e.

$$f(x) = \frac{6}{2}\sum_{j=1}^{2}\frac{b_j^{a_j}}{\Gamma(a_j)}(6x)^{a_j-1}e^{-6xb_j}$$

giving one mode and a plateau,

6. and a mixture of three gamma, $Gamma(a_j, b_j)$, $a_j = b_j^2$, $b_1 = 1.5$, $b_2 = 3$, and $b_3 = 6$ applied on $8x$ giving two bumps and one plateau.

The main mass is always in $[0, 1]$ with exponentially decreasing tails.

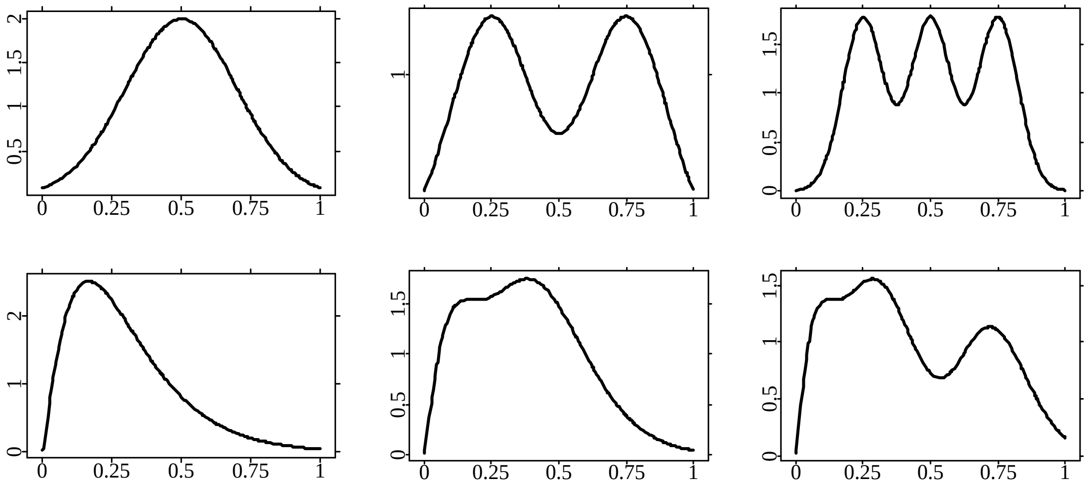

Figure 2: The six data generating densities: design 1 to 6 from the upper left to the lower right.

One may argue about what is a reasonable criteria on whose basis methods should be compared. Recall first that our objective is to minimize the integrated squared error (ISE) of the density estimate. This lead us to the following measures:

$$m_1 = \text{mean}(\text{ISE}(\widehat{h})), \quad m_2 = \text{std}(\text{ISE}(\widehat{h}))$$

$$m_3 = \text{mean}(\widehat{h} - \widetilde{h}), \quad m_4 = \text{std}(\widehat{h} - \widetilde{h}).$$

For brevity we have selected the results of the simulations when using Epanechnikov kernel and sample sizes $n = 50$, $n = 100$, and 200 as examples for moderate and large samples. The onesided crossvalidation presented here is the left-onesided. The original simulation study comprised more designs (DGPs), kernels, and samples sizes, but the findings were all the same. The criteria outcomes given in Tables 1 to 3 were calculated from 250 repetitions for each model and each sample size. From the simulation study one immediately notes that the unfeasible mixture mix1 is slightly better than the unfeasible plug-in method as our asymptotic theory indicated it would be. However, the improvement is insignificant in the overall picture and we can conclude that the second noise term in the plug-in method already is so small that it is irrelevant to improve more on it from a practical point of view. When considering feasible bandwidths, then first we note on the measure m_1 that the pragmatic average of classical crossvalidation and feasible plug-in beats each of its two components for almost every design and sample size. Therefore, from a practical point of view this simple average is much better than its two well known alternatives. We can get a hint of what is going on when looking on the bias m_3 and the volatility measures

m_2 and m_4. We see that the stability of the feasible plug-in method comes with the cost of a clear tendency to oversmoothing and that the unbiasness of crossvalidation comes with the cost of volatility. The simple average provides a good compromise between these two very different bandwidth selection methods. While mix3 works very well and probably is better than any other published bandwidth selector so far, we are still more ambitious in our search for good combinations of bandwidths. In that search onesided crossvalidation and do-validation turn out to become our bandwidth selectors of choice with do-validation coming out on top. Do-validation is only marginally better - measured on m_2 - than one-sided crossvalidation on most of the designs. But for the asymmetric density design 4 it is clearly better. The last thing to notice from this study is that classical crossvalidation is a crystal clear looser of this test. It is almost unbiased and that is good, but volatility just kills its overall performance. Therefore we suggest practitioners to leave classical crossvalidation and to start to use do-validation. Do-validation is just another crossvalidation technique, it is relatively simple to carry out, it is well defined without ambiguities and does not need complicated pilot estimation.

The conclusion the study is that do-validation is our preferred practical bandwidth selector.

Appendix

Proof of Theorem 1.

	Design 1								Design 2							
	h_0	PI	CV	OSCV	DO	Mix1	Mix2	Mix3	h_0	PI	CV	OSCV	DO	Mix1	Mix2	Mix3
$n = 50$																
m_1	.040	.047	.083	.051	.049	.039	.049	.049	.059	.115	.111	.106	.103	.059	.130	.083
m_2	.033	.033	.100	.043	.036	.032	.033	.037	.037	.023	.114	.033	.034	.039	.025	.032
m_3	.009	.047	-.011	.032	.034	.014	.059	.018	-.007	.117	.018	.092	.090	-.012	.138	.067
m_4	.035	.054	.095	.066	.064	.030	.053	.069	.034	.044	.078	.076	.072	.036	.047	.054
$n = 100$																
m_1	.026	.029	.049	.031	.030	.025	.031	.030	.036	.077	.063	.049	.049	.035	.091	.052
m_2	.019	.020	.059	.021	.020	.019	.021	.020	.020	.017	.055	.027	.026	.020	.017	.024
m_3	.003	.035	-.016	.019	.019	.007	.046	.010	.006	.098	.010	.035	.035	.005	.116	.054
m_4	.032	.047	.078	.057	.056	.029	.046	.058	.017	.024	.059	.048	.046	.015	.022	.039
$n = 200$																
m_1	.016	.017	.026	.018	.018	.015	.018	.018	.025	.049	.043	.034	.030	.024	.059	.033
m_2	.011	.012	.029	.014	.014	.011	.012	.013	.014	.013	.054	.041	.017	.014	.012	.016
m_3	.003	.029	-.005	.010	.011	.005	.036	.012	.005	.075	.000	.014	.015	.006	.091	.037
m_4	.026	.036	.064	.045	.044	.023	.045	.047	.014	.019	.044	.035	.031	.012	.017	.029

Table 1: Criteria m_1, m_2, m_3 and m_4 for designs 1 and 2.

For $L = K$ and $L = L_j$ $(j = 1, \ldots, J)$ we use the following notation. Define

$$\begin{aligned}
\Delta_L(h) &= \int \left(\widehat{f}_{L,h}(x) - f(x)\right)^2 dx \quad \text{(ISE)} \\
\mathrm{M}_L(h) &= \mathrm{E}\left[\Delta_L(h)\right] \quad \text{(MISE)} \\
\mathrm{D}_L(h) &= \Delta_L(h) - \mathrm{M}_L(h) \\
\delta_L(h) &= 2\int f(x)\widehat{f}_{L,h}(x)dx - 2n^{-1}\sum_{i=1}^{n} \widehat{f}_{L,h,-i}(X_i).
\end{aligned}$$

Here $\widehat{f}_{L,h}(x)$ denotes the kernel density estimator with bandwidth h and kernel L and $\widehat{f}_{L,h,-i}$ is the kernel density estimator obtained by leaving out the sample value X_i.

	Design 3								Design 4							
	h_0	PI	CV	OSCV	DO	Mix1	Mix2	Mix3	h_0	PI	CV	OSCV	DO	Mix1	Mix2	Mix3
$n = 50$																
m_1	.074	.158	.130	.156	.156	.075	.163	.126	.083	.130	.138	.114	.109	.083	.144	.105
m_2	.036	.011	.117	.015	.016	.039	.011	.024	.049	.063	.124	.061	.060	.049	.067	.058
m_3	-.001	.163	.036	.154	.154	-.008	.189	.099	.007	.095	.007	.063	.056	.007	.114	.051
m_4	.026	.026	.080	.039	.037	.028	.028	.047	.028	.054	.074	.060	.057	.026	.057	.057
$n = 100$																
m_1	.046	.142	.070	.115	.115	.046	.153	.091	.055	.087	.078	.072	.068	.055	.099	.066
m_2	.025	.008	.057	.042	.036	.025	.009	.019	.032	.042	.054	.038	.037	.032	.047	.035
m_3	-.002	.146	.007	.104	.106	-.004	.175	.076	.002	.075	-.006	.037	.030	.004	.092	.035
m_4	.012	.012	.042	.067	.059	.012	.012	.025	.024	.042	.050	.049	.045	.023	.046	.041
$n = 200$																
m_1	.030	.120	.042	.039	.038	.029	.136	.063	.034	.053	.049	.048	.040	.034	.062	.039
m_2	.013	.006	.032	.024	.021	.013	.008	.015	.019	.023	.046	.054	.021	.019	.026	.021
m_3	-.001	.127	.000	.018	.018	-.002	.154	.064	.006	.062	-.007	.026	.019	.008	.077	.027
m_4	.009	.010	.030	.034	.029	.008	.010	.019	.018	.027	.038	.042	.033	.018	.030	.029

Table 2: Criteria m_1, m_2, m_3 and m_4 for designs 3 and 4.

Put $c_{1,L} = \int L^2(u)du$, $c_{2,L} = \left(\int u^2 L(u)du\right)^2$. Define

$$\begin{aligned}
h_{L,0} &= \arg\min_h \mathrm{M}_L(h) \\
\widehat{h}_{L,0} &= \arg\min_h \Delta_L(h) \\
\widehat{h}_{L,c} &= \arg\min_h \left(\Delta_L(h) + \delta_L(h) - \int f(x)^2 dx\right) \quad \text{(CV-bandwidths).}
\end{aligned}$$

Under our conditions it holds that

$$h_{L,0} = \left(\frac{c_{1,L}}{c_{2,L} R_{f''}}\right)^{1/5} n^{-1/5} + o\left(n^{-3/10}\right).$$

Proceeding as in Hall and Marron (1987) one can show that

$$\begin{aligned}
\widehat{h}_{L,0} - h_{L,0} &= -\,\mathrm{M}''_L(h_{L,0})^{-1}\,\mathrm{D}'_L(h_{L,0}) + o_p\left(n^{-3/10}\right) \quad \text{(for } L = K, L_1, \ldots, L_J\text{),} \\
\widehat{h}_{L,c} - h_{L,0} &= -\,\mathrm{M}''_L(h_{L,0})^{-1}\left(\mathrm{D}'_L(h_{L,0}) + \delta'_L(h)\right) + o_p\left(n^{-3/10}\right) \quad \text{(for } L = L_1, \ldots, L_J\text{).}
\end{aligned}$$

	Design 5								Design 6							
	h_0	PI	CV	OSCV	DO	Mix1	Mix2	Mix3	h_0	PI	CV	OSCV	DO	Mix1	Mix2	Mix3
								$n = 50$								
m_1	.055	.063	.093	.064	.064	.054	.066	.061	.054	.073	.090	.070	.070	.053	.081	.063
m_2	.025	.023	.097	.025	.025	.025	.024	.027	.026	.020	.076	.027	.027	.026	.018	.026
m_3	.007	.078	.008	.069	.064	.007	.092	.043	.006	.104	.015	.074	.071	.004	.122	.060
m_4	.048	.057	.099	.064	.067	.045	.057	.073	.036	.036	.093	.065	.066	.033	.033	.061
								$n = 100$								
m_1	.037	.046	.055	.045	.044	.037	.049	.040	.039	.056	.058	.049	.048	.038	.062	.045
m_2	.015	.013	.045	.015	.015	.015	.014	.015	.016	.015	.045	.018	.019	.016	.014	.017
m_3	.000	.074	-.002	.059	.051	.000	.090	.036	-.003	.096	.005	.057	.053	-.004	.115	.051
m_4	.038	.045	.072	.055	.056	.039	.047	.052	.028	.028	.068	.051	.052	.026	.026	.044
								$n = 200$								
m_1	.024	.034	.033	.032	.029	.024	.037	.027	.025	.042	.035	.033	.031	.025	.048	.031
m_2	.010	.008	.020	.010	.010	.010	.009	.010	.010	.009	.031	.012	.012	.010	.009	.010
m_3	.003	.077	.000	.055	.045	.003	.094	.039	-.002	.093	.001	.048	.041	-.003	.113	.047
m_4	.027	.032	.057	.047	.044	.027	.033	.040	.023	.023	.050	.044	.041	.022	.022	.034

Table 3: Criteria m_1, m_2, m_3 and m_4 for designs 5 and 6.

For the proof of these statements it can be checked that it is not needed the L is symmetric and continuous at the point 0. Proceeding as in Hall and Marron (1987) one can show that for $L = K, L_1, \ldots, L_J$ with $h = h_{L,0}$:

$$\mathrm{D}'_L(h) = n^{-2} \sum_{i<j} W_{L,i,j} + n^{-1} \sum_i W_{L,i} + o_p\left(n^{-7/10}\right)$$

with $W^*_{L,i,j} = -h^{-2} H_L \left(\frac{X_i - X_j}{h} \right)$,

$$H_L(u) = 2\int L(u+v)L(v)dv + 2\int L(-u+v)L(v)dv + 2\int L(u+v)vL'(v)dv + 2\int L(-u+v)vL'(v)dv,$$

$$\begin{aligned} W_{L,i,j} &= W^*_{L,i,j} - \mathrm{E}\left[W^*_{L,i,j}|X_i\right] - \mathrm{E}\left[W^*_{L,i,j}|X_j\right] + \mathrm{E}\left[W^*_{L,i,j}\right] \\ W^*_{L,i} &= 2h\sqrt{c_{2,L}} f''(X_i) \\ W_{L,i} &= W^*_{L,i} - \mathrm{E}\left[W^*_{L,i}\right] \end{aligned}$$

Furthermore, we have for $h = h_{L,0}$ and $L = L_1, \ldots, L_J$

$$\delta'_L(h) = n^{-2}\sum_{i<j} V_{L,i,j} - n^{-1}\sum_i W_{L,i} + o_p\left(n^{-7/10}\right)$$

with $V^*_{L,i,j} = h^{-2} M_L\left(\frac{X_i - X_j}{h}\right)$, $M_L(u) = 2\left[L(u) + uL'(u) + L(-u) - uL'(-u)\right].$

We now use

$$\begin{aligned} h_{K,0} &= \left(\frac{c_{1,K}}{c_{2,K}}\frac{c_{2,L_j}}{c_{1,L_j}}\right)^{1/5} h_{L_j,0} + o\left(n^{-3/10}\right) \\ &= \left(\frac{c_1}{c_2}\frac{c_{2,j}}{c_{1,j}}\right)^{1/5} h_{L_j,0} + o\left(n^{-3/10}\right). \end{aligned}$$

This gives together with the above expansions:

$$\begin{aligned} \widehat{h} - \widetilde{h} &= \sum_{j=1}^{J} w_j \left(\frac{c_1}{c_2}\frac{c_{2,j}}{c_{1,j}}\right)^{1/5} (\widehat{h}_{L_j,c} - h_{L_j,0}) - (\widetilde{h} - h_{K,0}) + o\left(n^{-3/10}\right) \\ &= \sum_{j=1}^{J} w_j M''_{L_j}(h_{L_j,0})^{-1} \left(\frac{c_1}{c_2}\frac{c_{2,j}}{c_{1,j}}\right)^{1/5} \left[-n^{-2}\sum_{i<k}\left(W_{L_j,i,k} + V_{L_j,i,k}\right)\right] \\ &\quad - M''_K(h_0)^{-1}\left[-n^{-2}\sum_{i<k} W_{K,i,k} - n^{-1}\sum_i W_{K,i}\right] + o_p\left(n^{-3/10}\right) \end{aligned}$$

Now

$$\begin{aligned} M''_L(h_{L,0}) &= 2\frac{c_{1,L}}{nh^3_{L,0}} + 3c_{2,L}R_{f''}h^2_{L,0} + o\left(n^{-2/5}\right) \\ &= n^{-2/5} 5 c_{1,L}^{2/5} R_{f''}^{3/5} c_{2,L}^{3/5} + o\left(n^{-3/5}\right) \end{aligned}$$

With the above expansion this gives

$$\begin{aligned} \widehat{h} - \widetilde{h} &= M''_K(h_0)^{-1} n^{-1}\sum_i W_{K,i} \\ &\quad + M''_K(h_0)^{-1} n^{-2}\sum_{i<k} Z_{ik} + o_p\left(n^{-3/10}\right) \end{aligned}$$

with $Z_{ik} = Z^*_{ik} - \mathrm{E}\left[Z^*_{ik}|X_i\right] - \mathrm{E}\left[Z^*_{ik}|X_j\right] + \mathrm{E}\left[Z^*_{ik}\right]$,

$$\begin{aligned} Z^*_{ik} &= -h_{0,K}^{-2}\left[H_K\left(\frac{X_i - X_k}{h_0}\right)\right. \\ &\quad \left. - \sum_{j=1}^{J} w_j \left(\frac{c_1}{c_{1,j}}\right)\left(H_{L_j}\left(\frac{X_i - X_k}{h_{0,L_j}}\right) + M_{L_j}\left(\frac{X_i - X_k}{h_{0,L_j}}\right)\right)\right] \end{aligned}$$

The variance of the asymptotic expansion of $\widehat{h} - \widetilde{h}$ can be easily calculated. Furthermore, using a central limit theorem for U-statistics (e.g. Hall, 1984) one gets the asymptotic result for $\widehat{h} - \widetilde{h}$ in our theorem. The second statement of the theorem can be proved similarly.

References

Ahmad, I.A. and Ran, I.S., 2004, Data based bandwidth selection in kernel density estimation with parametric start via kernel contrasts, *Journal of Nonparametric Statistics.* **16**, 841–877.

Bowman, A., 1984, An alternative method of crossvalidation for the smoothing of density estimates. *Biometrika*, **71**, 353–360.

Chacón, J.E., Montanero, J. and Nogales, A.G., 2008, Bootstrap bandwidth selection using an h-dependent pilot bandwidth. *Scandinavian Journal of Statistics*, **35**, 139–157.

Chaudhuri, P. and Marron, J.S., 1999, SiZer for Exploration of Structures in Curves. *Journal of the American Statistical Association*, **94**, 807–823.

Cheng, M.Y., 1997a, Boundary-aware estimators of integrated squared density derivatives. *Journal of the Royal Statistical Society Ser. B*, **50**, 191–203.

Cheng, M.Y., 1997b, A bandwidth selector for local linear density estimators. *The Annals of Statistics*, **25**, 1001–1013.

Chiu, S.T., 1991, Bandwidth selection for kernel density estimation. *The Annals of Statistics*, **19**, 1883–1905.

Godtliebsen, F.; Marron, J.S. and Chaudhuri, P., 2002, Significance in Scale Space for Bivariate Density Estimation. *Journal of Computational and Graphical Statistics*, **11**, 1–21.

Härdle, W., Hall, P. and Marron, J.S., 1988, How far are automatically chosen regression smoothing parameters from their optimum?. *Journal of the American Statistical Association*, **83**, 86–99.

Hall, P., 1984, Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators. *Journal of the Multivariate Analysis*, **14**, 1–16.

Hall, P. and Johnstone, I., 1992, Empirical Functionals and Efficient Smoothing Parameter Selection. *Journal of the Royal Statistical Society B*, **54** (2), 475–530.

Hall, P. and Marron, J.S., 1987, Extent to which Least-Squares Cross-Validation Minimises Integrated Square Error in Nonparametric Density Estimation. *Probability Theory and Related Fields*, **74**, 567–581.

Hall, P., Marron, J.S. and Park, B., 1992, Smoothed crossvalidation. *Probability Theory and Related Fields*, **92**, 1–20.

Hanning, J. and Marron, J.S., 2006, Advanced Distribution Theory for SiZer. *Journal of the American Statistical Association*, **101**, 484–499.

Hart, J.D., and Li., 2005, Robustness of one-sided cross-validation to autocorrelation. *Journal of Multivariate Statistics*, **92**, 77–96.

Hart, J.D. and Yi, S., 1998, One-Sided Cross-Validation. *Journal of the American Statistical Association*, **93**, 620–631.

Jones, M.C., 1993, Simple boundary correction in kernel density estimation. *Statistics and Computing*, **3**, 135–146.

Loader, C.R., 1999, Bandwidth selection: classical or plug-in. *Ann. Statist.*, **27**, 415–438.

Martínez-Miranda, M.D., Nielsen, J. and Sperlich, S., 2009, One sided cross-validation for density estimation with an application to operational risk. *In "Operational Risk Towards Basel III: Best Practices and Issues in Modelling. Management and Regulation, ed. G.N. Gregoriou; John Wiley and Sons, Hoboken, New Jersey.*

Rudemo, M., 1982, Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, **9**, 65–78.

Savchuk, O.Y., Hart, J.D., and Sheather S.J., 2010a, Indirect cross-validation for Density Estimation. Submitted to *Journal of the American Statistical Association*, available at `http://arxiv.org/abs/0812.0051`.

Savchuk, O.Y., Hart, J.D., and Sheather S.J., 2010b, An empirical study of indirect cross-validation for Density Estimation. *IMS Lecture Notes - Festschrift for Tom Hettmansperger.*

Scott, D.W. and Terrell, G.R., 1987, Biased and unbiased crossvalidation in density estimation. *Journal of the American Statistical Association*, **82**, 1131–1146.

Sheather, S.J. and Jones, M.C., 1991, A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society*, Ser. B, **53**, 683–690.

Żychaluk, K. and Patil, P.N., 2008, A cross-validation method for data with ties in kernel density estimation. *Annals of the Institute of Statistical Mathematics*, **60**, 21–44.