One-Sided Cross-Validation for Nonsmooth Density Functions

Olga Y. Savchuk

Abstract

One-sided cross-validation (OSCV) is a bandwidth selection method initially introduced by Hart and Yi (1998) in the context of smooth regression functions. Martínez-Miranda et al. (2009) developed a version of OSCV for smooth density functions. This article extends the method for nonsmooth densities. It also introduces the fully robust OSCV modification that produces consistent OSCV bandwidths for both smooth and nonsmooth cases. Practical implementations of the OSCV method for smooth and nonsmooth densities are discussed. One of the considered cross-validation kernels has potential for improving the OSCV method's implementation in the regression context.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability density function f. The kernel density estimator of f is computed as

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

where $h > 0$ is a smoothing parameter that is usually called the bandwidth, and the kernel K is assumed to be of the second order, which means that $\int_{-\infty}^{\infty} K(u)\, du = 1$, $\int_{-\infty}^{\infty} uK(u)\, du = 0$, and $\int_{-\infty}^{\infty} u^2 K(u)\, du < \infty$. Most frequent choices of K include the Gaussian, Epanechnikov, and quartic kernels (see Wand and Jones (1995)).

The two most commonly used measures of performance of $\hat{f}_{h,K}$ are the integrated squared error (ISE) and the mean integrated squared error (MISE) defined as

$$\text{ISE}_K(h) = \int_{-\infty}^{\infty} \left(\hat{f}_{h,K}(x) - f(x)\right)^2 dx,$$

$$\text{MISE}_K(h) = \text{E}\left(\text{ISE}_K(h)\right).$$

Let $\hat{h}_0$ and h_0 denote the minimizers of the ISE_K and MISE_K functions, respectively.

Both bandwidths $\hat{h}_0$ and h_0 are unavailable for practical use since their computation requires knowing f. There exist many data-driven bandwidth selection techniques (see the survey of Jones et al. (1996)). Some of the most popular bandwidth selectors are the plug-in rule of Sheather and Jones (1991) and the least squares cross-validation (LSCV) method proposed independently by Rudemo (1982) and Bowman (1984).

The cross-validation method is quite popular among practitioners because of its simplicity. Moreover, it requires fewer assumptions on f compared to the plug-in method. Nevertheless, the method is criticized because of producing too variable bandwidths and selecting the trivial bandwidths for the data sets that contain substantial amount of tied observations (see Silverman (1986) and Chiu (1991)). A well known LSCV paradox consists in the method's improved performance on the harder estimation problems (see Loader (1999)). A couple of successful modifications of the LSCV method that take advantage of this paradox are the one-sided cross-validation (OSCV) method proposed by Martínez-Miranda et al. (2009) and the indirect cross-validation (ICV) method of Savchuk et al. (2010). Both methods are supported by the corresponding R packages (see Savchuk (2017a) and Savchuk (2017b)).

The OSCV method is originally introduced in the regression context (see Hart and Yi (1998)). Martínez-Miranda et al. (2009) extended the method to the case of smooth densities. A density function f is referred to as *smooth* if it is twice continuously differentiable, whereas it is called *nonsmooth* if it is continuous but has finitely many simple discontinuity points in its first derivative. The OSCV method in the smooth case is shown to greatly stabilize the bandwidth distribution (see Martínez-Miranda et al. (2009)). This article extends the OSCV method to the case of nonsmooth density functions.

Martínez-Miranda et al. (2009) introduced the left-sided and right-sided OSCV versions based on the so-called left-sided and right-sided kernels, respectively. Both one-sided kernels are obtained by multiplying a benchmark two-sided kernel by a linear function and restricting the support of the one-sided kernel to either $(-\infty, 0]$ (left-sided case) or $[0, \infty)$ (right-sided case). In this article we restrict our attention on the right-sided OSCV version.

The right-sided kernel L based on the benchmark two-sided kernel H is computed as

$$L(u) = \frac{\int_0^\infty t^2 H(t)\, dt - u \int_0^\infty t H(t)\, dt}{\int_0^\infty H(t)\, dt \int_0^\infty t^2 H(t)\, dt - \left(\int_0^\infty t H(t)\, dt\right)^2} H(u) I_{[0,\infty)}(u). \tag{1}$$

It follows that L is of the second order. Generally, the benchmark kernel H is different from the kernel K used to compute $\hat{f}_{h,K}$. Martínez-Miranda et al. (2009) implemented OSCV based on $H = K = K_E$, where K_E denotes the Epanechnikov kernel.

The OSCV function based on a one-sided kernel L is defined as

$$\text{OSCV}_L(b) = R(\hat{f}_{b,L}) - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{b,L}^{-i}(X_i), \tag{2}$$

where the definition of $R(\cdot)$ is given in the Appendix. In the above expression $\hat{f}_{b,L}$ is the density estimator based on the kernel L and the bandwidth b, whereas $\hat{f}_{b,L}^{-i}(X_i)$ is its leave-one-out modification that is computed from all data points except X_i. The above version of the OSCV function mimics the traditional definition of the CV function of Rudemo (1982) and Bowman (1984) and slightly differs from the function used in Martínez-Miranda et al. (2009) and the follow-up articles (see Mammen et al. (2011) and Mammen et al. (2014)). Indeed, in the OSCV function of Martínez-Miranda et al. (2009), the estimator $\hat{f}_{b,L}^{-i}$ under the sum is replaced by $\hat{f}_{b,L}$. This is justified by assuming $L(0) = 0$. We find this assumption rather restrictive since the one-sided versions of the most frequently used kernels do not possess this property. Since the case $L(0) \neq 0$ does not substantially complicate the OSCV implementation, we proceed by using (2). Let $\hat{b}$ denote the minimizer of (2).

Martínez-Miranda et al. (2009) defined the OSCV bandwidth in the smooth case as $\hat{h}_{OSCV} = C \cdot \hat{b}$, where

$$C = \left(\frac{R(K)}{R(L)} \cdot \frac{\mu_2^2(L)}{\mu_2^2(K)} \right)^{1/5}. \tag{3}$$

The functionals $R(\cdot)$ and $\mu_2(\cdot)$ are defined in the Appendix. The OSCV bandwidth $\hat{h}_{OSCV}$ is consistent for the MISE optimal bandwidth h_0, that is $\hat{h}_{OSCV} \xrightarrow{p} h_0$.

2 OSCV for nonsmooth density functions

In this section we extend the OSCV algorithm to the case of a nonsmooth density f that has simple discontinuities in its first derivative at the points $\{x^{(t)}\}$, $t = 1, 2, \ldots, k$. The extension is based on the asymptotic expansion of MISE of the kernel density estimator in the nonsmooth case (see Cline and Hart (1991) and van Es (1992)) that has the following form:

$$\text{MISE}_K(h) = \text{AMISE}_K^*(h) + o\left(h^3 + \frac{1}{nh} \right),$$

Kernel K	C	C^*	E_C
Epanechnikov	0.5371	0.5019	7.01
quartic	0.5573	0.5206	7.05
Gaussian	0.6168	0.5730	7.64

Table 1: Rescaling constants for the most frequently used kernels.

where

$$\text{AMISE}_K^*(h) = \frac{R(K)}{nh} + h^3 B(K) \sum_{i=1}^{k} \left(f'\left(x^{(t)}+\right) - f'\left(x^{(t)}-\right)\right)^2,$$

where $B(\cdot)$ is defined in the Appendix. The above expression yields the following asymptotically optimal bandwidth h_n^*:

$$h_n^* = \left(\frac{R(K)}{3B(K)\sum_{i=1}^{k}\left(f'\left(x^{(t)}+\right) - f'\left(x^{(t)}-\right)\right)^2}\right)^{1/4} n^{-1/4}.$$

The asymptotic expansion of $\text{MISE}_L(b)$, the MISE function for $\hat{f}_{b,L}$, has the same form as that of $\text{MISE}_K(h)$ with K and h being replaced by L and b. Let b_n^* denote the asymptotically optimal bandwidth for $\text{MISE}_L(b)$. It then follows that

$$\frac{h_n^*}{b_n^*} = C^* = \left(\frac{R(K)}{B(K)} \cdot \frac{B(L)}{R(L)}\right)^{1/4}. \tag{4}$$

This motivates defining the OSCV bandwidth in the nonsmooth case as $\hat{h}_{OSCV} = C^* \cdot \hat{b}$.

In both smooth and nonsmooth cases the OSCV bandwidth is defined by multiplying $\hat{b}$, the minimizer of the OSCV function (2), by a rescaling constant. The constant C (3) is used in the smooth case, whereas the constant C^* (4) is used in the nonsmooth case.

It is remarkable that the expressions for C and C^* are identical to those in the OSCV implementation for regression functions (see Savchuk et al. (2013) and Savchuk et al. (2016)). This follows from similarity of the corresponding asymptotic expansions of MISE of the kernel density estimator and the mean average squared error (MASE) of the local linear estimator. The values of C and C^* in the case $H = K$ for the most frequently used kernels K and their one-sided counterparts L are given in Table 1. The quantity E_C that appears in the last column of Table 1 is defined by

$$E_C = \left(\frac{C}{C^*} - 1\right) \cdot 100\%.$$

4

E_C assesses the magnitude and direction of the asymptotic relative bandwidth bias introduced by using C instead of C^* in the nonsmooth case.

In this article we set $K = \phi$, where ϕ denotes the Gaussian kernel, and assess how different one-sided kernels change the theoretical properties and practical performance of the OSCV method. In the example's section below we investigate performances of L_G and L_E, the one-sided Gaussian and Epanechnikov kernels, respectively, for smooth and nonsmooth densities. The one-sided kernels discussed in Section 4 increase the method's resistance against potential nonsmoothness of f, at least theoretically.

3 Examples

Example 1. This example shows that in the nonsmooth case the actual bias of $\hat{h}_{OSCV}$ based on the smooth constant C tends to be even greater than the corresponding value of E_C from Table 1.

We generated 1000 data sets of size $n = 500$ from the density f^* shown in Figure 1 **(a)**. This density can be found in the R package `OSCV` that allows generating random samples from this density and computing the f^*-based ISE functions (see Savchuk (2017b)).
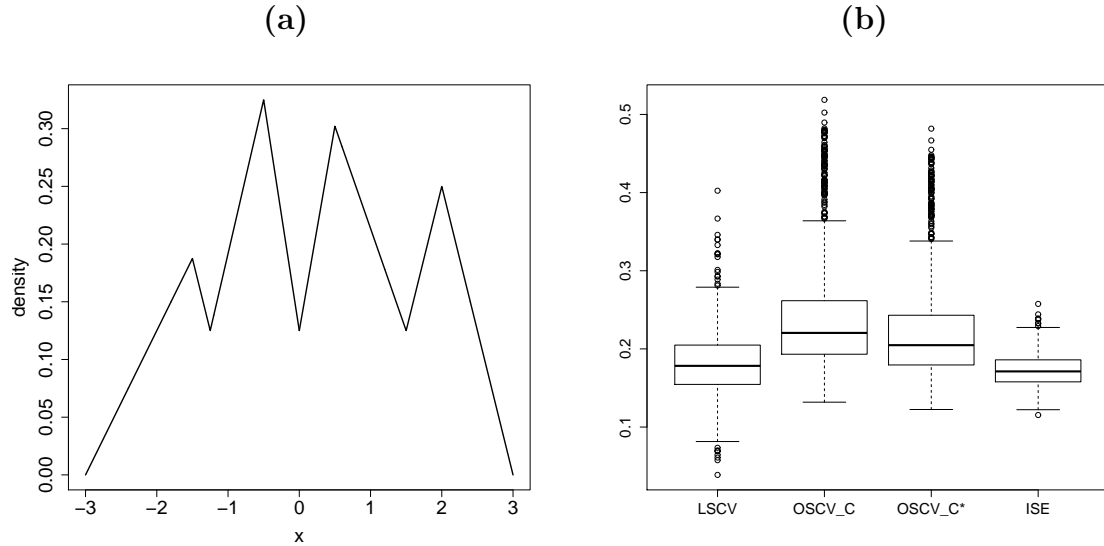
(a)   **(b)**

Figure 1: **(a)** The density f^* with 7 cusps. **(b)** Boxplots for LSCV, ISE and OSCV versions based on L_G that use C and C^*.

Figure 1 **(b)** displays the boxplots of the bandwidths selected by different methods. Two

Method	OSCV based on C	OSCV based on C^*	LSCV
Δ_B	28.76	19.61	4.14
Δ_{ISE}	13.64	10.00	6.57

Table 2: Measures of performance computed from 1000 replications in the case of f^* and $n = 500$.

boxplots in the middle show the distributions of OSCV bandwidths selected by L_G and based on the rescaling constants C and C^* from Table 1. Also shown are the boxplots of the LSCV bandwidths and the ISE-optimal bandwidths $\hat{h}_0$.

We used the following measures of performance of a data-driven bandwidth $\hat{h}$:

$$
\Delta_B = \frac{\hat{M}(\hat{h}) - \hat{M}(\hat{h}_0)}{\hat{M}(\hat{h}_0)} \cdot 100\%,
$$
$$
\Delta_{ISE} = \hat{M}\left(\frac{\text{ISE}(\hat{h}) - \text{ISE}(\hat{h}_0)}{\text{ISE}(\hat{h}_0)}\right) \cdot 100\%,
$$
(5)

where $\hat{M}(Y)$ denotes the median of a random variable Y computed over 1000 replications. In the nonsmooth case, the value of Δ_{ISE} is an empirical analog of E_C for $\hat{h}_{OSCV}$ based on the smooth constant C.

Table 2 shows the values of Δ_B and Δ_{ISE} from our simulations for three considered data-driven methods. The value of $\Delta_{ISE} = 13.64$ for OSCV based on C is much greater than $E_C = 7.64$ from Table 1. This is explained by the fact that the OSCV method based on L_G produces greatly variable bandwidths that tend to be inappropriately large even in the case when the nonsmooth constant C^* is used. Figure 2 (a) shows a typical L_G-based curve computed for a random sample generated from f^* at $n = 500$. The horizontal scale of the graph is changed such that the minimum is observed at the C^*-based OSCV bandwidth of 0.3179. The ISE-optimal bandwidth for these data is $\hat{h}_0 = 0.1673$.

The LSCV method outperforms both OSCV versions in terms of the measures (5). This is consistent with the aforementioned LSCV paradox since the density f^* is relatively hard to estimate. Figure 3 illustrates this by showing a density estimate based on the LSCV bandwidth for a random sample generated from f^* at $n = 300$. The ISE-optimal bandwidth for this data set is $\hat{h}_0 = 0.1884$.

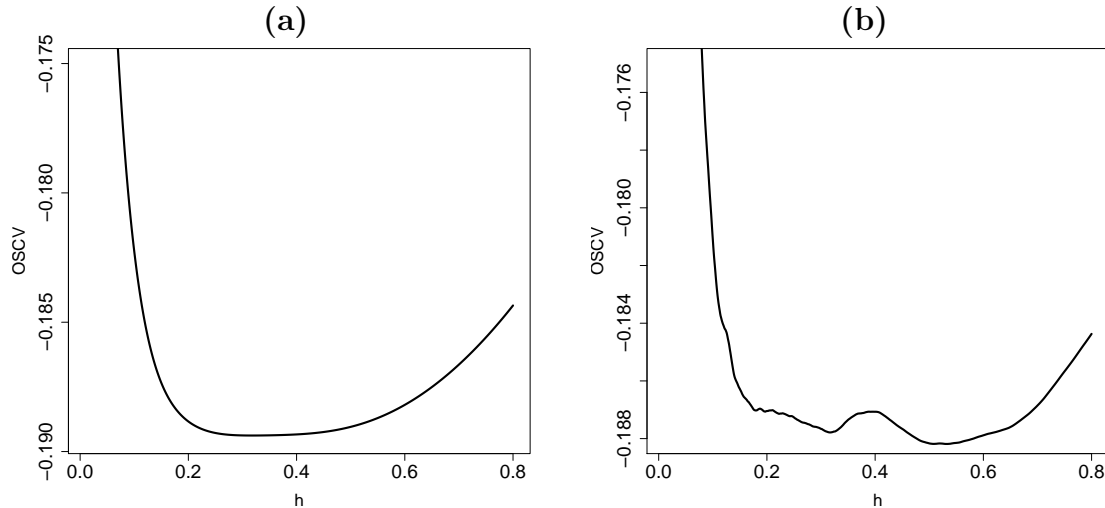The original implementation of the OSCV method of Martínez-Miranda et al. (2009) is

Figure 2: OSCV curves based on **(a)** L_G and **(b)** L_E for a random sample generated from f^* at $n = 500$.

based on L_E, the one-sided Epanechnikov kernel. Figure 2 **(b)** shows the L_E-based OSCV curve computed for the same realization that yields the curve in Figure 2 **(a)**. The horizontal scale of the graph in Figure 2 **(b)** is changed such that its minimizer is to be plugged-in to the Gaussian density estimator without additional rescaling. However, the curve's minimizer of 0.4940 appears to be too large and results in an oversmoothed density estimate.

The curve in Figure 2 **(b)** is unacceptably wiggly. We generated many other samples from f^* and found that the most of the corresponding L_E-based OSCV curves are inappropriately nonregular whereas the L_G-based OSCV curves are usually smooth and have one local minimum. This can be seen by experimenting with the R functions `OSCV_Epan_dens` and `OSCV_Gauss_dens` from the R package `OSCV` (see Savchuk (2017b)).

The problem of producing insufficiently smooth OSCV curves by L_E persists in the smooth case, as the Example 2 illustrates. Roughness of the OSCV curves based on the Epanechnikov kernel was also noted in the regression context (see Savchuk and Hart (2017)).

Example 2. A random sample of size $n = 100$ was generated from the standard normal density. Figure 4 shows the corresponding L_G-based and L_E-based OSCV curves. The curves' horizontal scales are adjusted such that their minimizers are to be used to compute the Gaussian density estimates without further rescaling. In fact, Figure 4 **(a)** and **(b)** shows the curves $\text{OSCV}_{L_G}(b/C_G)$ and $\text{OSCV}_{L_E}(b/C_E)$, where the rescaling constants C_G and C_E are obtained by using (3) with $K = \phi$ and L corresponding to either L_G or L_E, respectively.
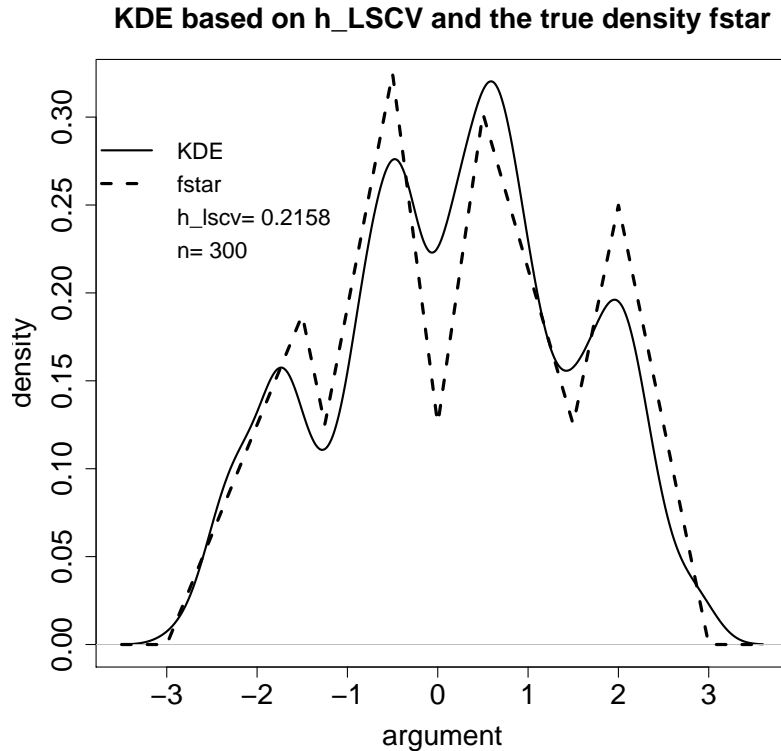
7

KDE based on h_LSCV and the true density fstar

Figure 3: Density estimate based on the LSCV bandwidth for a realization generated from f^* at $n = 300$.

The bandwidths selected by L_E and L_G are 0.4512 and 0.4714, correspondingly, whereas the ISE-optimal bandwidth $\hat{h}_0 = 0.4423$. The L_E-based OSCV curve is inappropriately wiggly. In particular, it has two local minima of about the same size located at 0.3268 and 0.4512. Just by luck the latter local minimum appears to be somewhat smaller.

For all other smooth densities that we considered in our numerous simulation experiments, the L_G-based OSCV curves were usually smoother than the corresponding L_E-based curves. Thus, L_G appears to be a better candidate than L_E for the OSCV method's implementation in the smooth case.

The problem of occasionally producing rough criterion curves by L_E persists in the case of real data. This is illustrated by Example 3.

Example 3. We used the famous data set of size $n = 272$ on the eruption duration of the Old Faithful geyser in Yellowstone National Park that can be found in Härdle (1991). The corresponding L_E- and L_G-based OSCV curves are plotted in Figure 5 **(a)** and **(b)**,
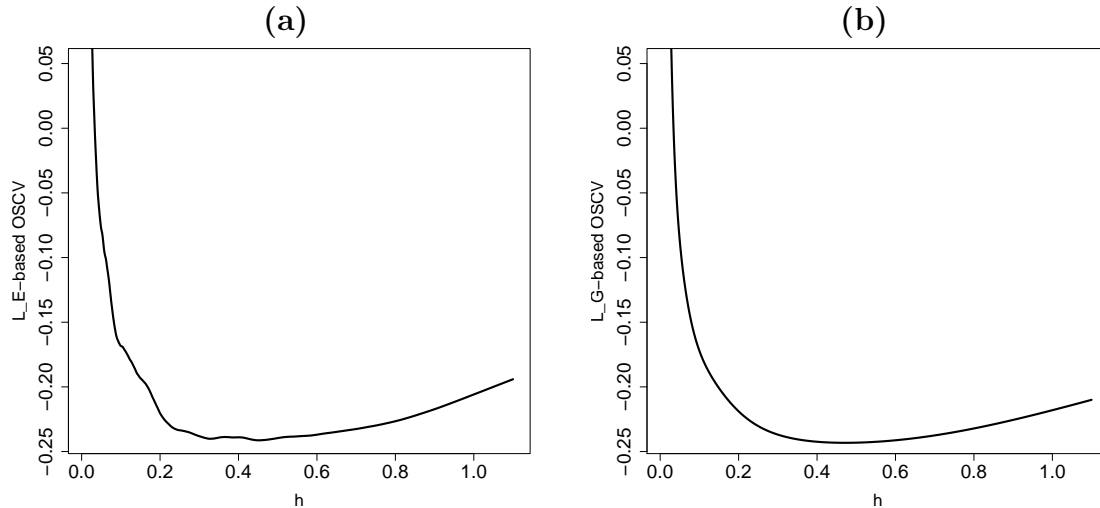
Figure 4: OSCV curves based on **(a)** L_E and **(b)** L_G for a random sample generated from the standard normal distribution at $n = 100$.

respectively. The horizontal scales of the graphs are not adjusted, so additional rescaling of the curves' minimizers is needed before they can be plugged into the Gaussian density estimator. The L_E -based graph shown in Figure 5 **(a)** can be reproduced in R by using the code from the Examples section of the R package `OSCV`.

Obviously, the L_E-based curve for the eruption duration data is unacceptably wiggly even for the values of b near the curve's minimizer. To the contrary, the L_G-based curve is perfectly regular. Assuming that the distribution of the eruption duration is smooth, the L_G-based OSCV bandwidth is found as $\hat{h}_{OSCV} = 0.6168 \cdot 0.2037 = 0.1256$. For comparison, the LSCV and Sheather-Jones plug-in bandwidths are 0.1021 and 0.1395, respectively.

4 Fully Robust OSCV for density functions

The fully robust OSCV method proposed in the regression context (see Savchuk et al. (2013) and Savchuk et al. (2016)) can be adapted for the density estimation setting. The underlying idea is that the OSCV function (2) is computed based on a so-called *robust* one-sided kernel L that has equal smooth and nonsmooth rescaling constants, C and C^*, respectively. As a consequence, the OSCV bandwidth selected by a robust kernel is consistent for the MISE optimal bandwidth h_0 in both smooth and nonsmooth cases.

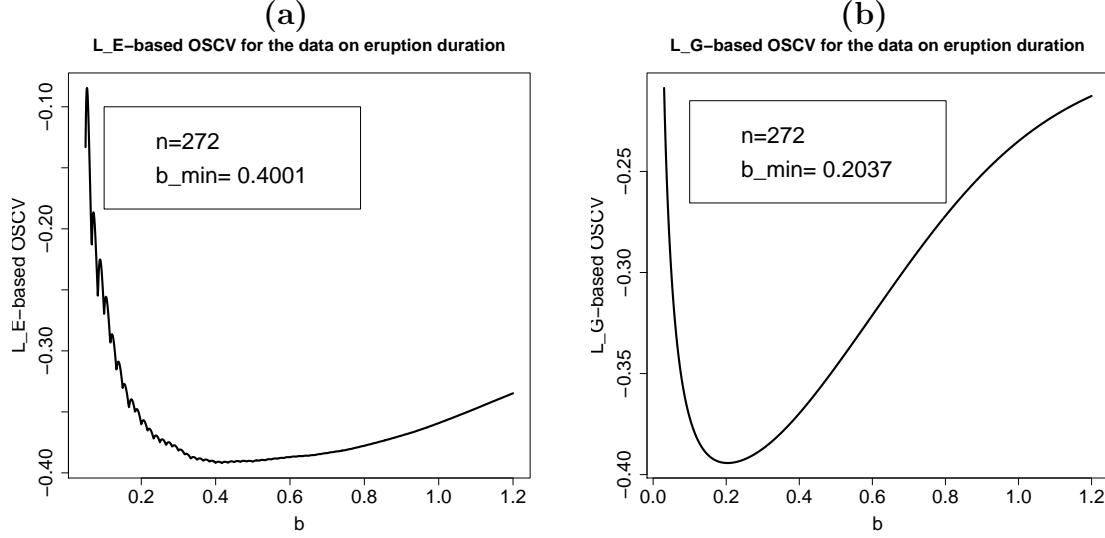Since the density f^* introduced in Example 1 is fairly nonsmooth, we used it for initial

Figure 5: OSCV curves based on **(a)** L_E and **(b)** L_G for the data on eruption duration of the Old Faithful geyser.

evaluation of performances of the robust and *almost* robust one-sided kernels that we managed to find. A kernel L is called almost robust if it has $E_C < 5\%$ (see Savchuk and Hart (2017)). The same approach was previously used in the regression fully robust OSCV version where the performances of the robust and almost robust kernels were initially judged on the regression function r_3 that has six cusps. Thus, we concentrated on searching for a one-sided robust kernel L that, at least, outperforms LSCV in the case of f^* and produces reasonably smooth OSCV curves.

The robust kernel used in the fully robust OSCV implementation in the regression context (see Savchuk and Hart (2017)) is also robust in the density estimation setting. This kernel is a member of the following family:

$$L_I(u; \alpha, \sigma) = \frac{a + bu}{c} H_I(u) I_{[0,\infty)}(u), \qquad \alpha \in \mathbb{R}, \ \sigma > 0, \tag{6}$$

where $H_I(u) = (1 + \alpha)\phi(u) - \alpha\phi(u/\sigma)/\sigma$ is a two-sided counterpart of L_I, and

$$a = 2\pi(1 + \alpha - \alpha\sigma^2);$$
$$b = -2\sqrt{2\pi}(1 + \alpha - \alpha\sigma);$$
$$c = \pi(1 + \alpha - \alpha\sigma^2) - 2(1 + \alpha - \alpha\sigma)^2.$$

The kernel has $(\alpha, \sigma) = (16.8954588, 1.01)$ and is plotted in Figure 6 **(a)**. It is worth to mention that the one-sided Gaussian kernel L_G is obtained from (6) by either setting $\alpha = 0$ or $\sigma = 1$.

(a) Kernel with $(\alpha, \sigma) = (16.8954588, 1.01)$

(b) Kernel with $(\alpha, \sigma) = (0.4275, 10)$

(c) Kernel with $(\alpha, \sigma) = (4, 0.8)$
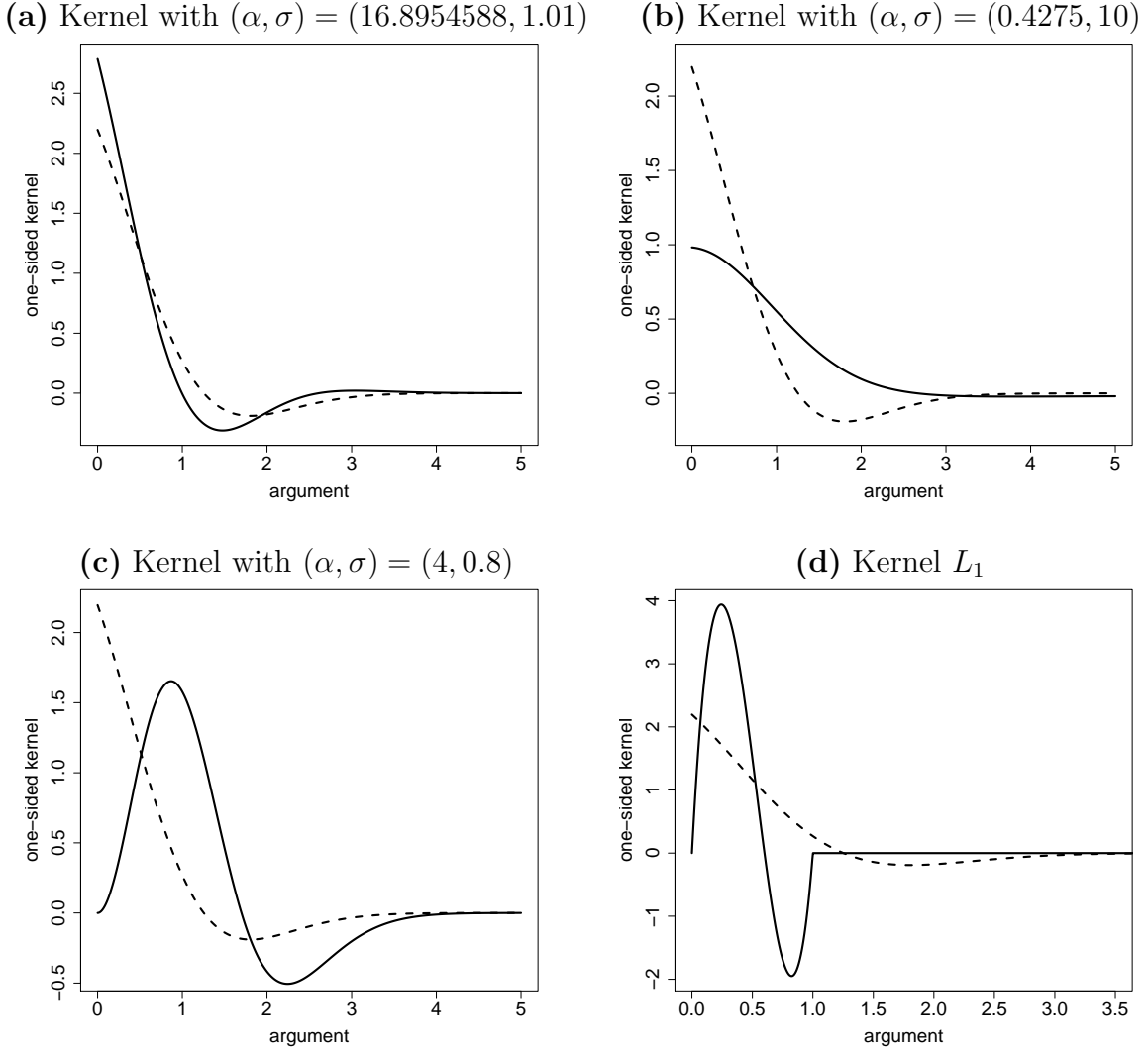
(d) Kernel L_1

Figure 6: The solid curves show the one-sided kernels, whereas the dashed curves show L_G.

Despite the fact that the kernel with $(\alpha, \sigma) = (16.8954588, 1.01)$ works well in the regression context (see Savchuk and Hart (2017)), its performance in the density estimation framework is unsatisfactory. Our inspection of the L_I-based OSCV curves for a sequence of random samples generated from f^* revealed that the considerable part of them has two local minima with the largest one being inappropriately large. However, the more serious problem with L_I is that it frequently produces the OSCV curves that tend to $-\infty$ as $h \to 0$. Silverman (1986) and Chiu (1991) argued that the LSCV method experiences this type of problem even for such frequently used kernels as Epanechnikov, quartic, and Gaussian.

Inappropriate performance of the kernel with $(\alpha, \sigma) = (16.8954588, 1.01)$ on f^* stimulated further search for the robust kernels. First of all, we inspected performances of two other

robust kernels mentioned in Savchuk and Hart (2017). Both these kernels are members of the family (6) with $\sigma = 10$ and the values of α equal to 0.4275 and 0.9821. One of the kernels (with $\alpha = 0.4275$) is plotted in Figure 6 **(b)**. The other kernel (with $\alpha = 0.9821$) has a similar shape and performance. Unfortunately, both kernels are found to produce highly variable and biased upwards bandwidth distributions, at least in the case of f^*.

In our next attempt, we considered another member of (6) with $(\alpha, \sigma) = (4, 0.8)$ that is plotted in Figure 6 **(c)**. This kernel is almost robust with $E_C = 1.17\%$. It has a quite different shape compared to the one-sided kernels considered so far, but, unfortunately, performs even worse than them. Indeed, for all inspected realizations from f^* and the $N(0,1)$ density it produced the OSCV curves that tended to $-\infty$ as $h \to 0$. It is remarkable that the kernel with $(\alpha, \sigma) = (4, 0.8)$ is equal to zero at the origin. This implies that H_I, the two-sided counterpart of this kernel, is nonnegative and bimodal (see Savchuk et al. (2010)). According to Savchuk et al. (2013) and Savchuk and Hart (2017) such a kernel has potential for OSCV implementation in the regression context.

Further experimenting with the kernels from the family (6) is possible by using the R function `OSCV_LI_dens` from the R package `OSCV`. Indeed, many other robust kernels can be found in (6). It is entirely possible that there exists one that performs better than the kernels discussed above.

Three other almost robust kernels considered below are not the members of (6) but originate from the dissertation of Yi (1996). These kernels, denoted by L_1 L_2, and L_3, are defined below. Figure 6 **(d)** shows L_1. The graphs of the other two kernels are not included since they are fairly similar in shape to L_1.

$$L_1(u) = 6u(1 - u)(6 - 10u)I_{[0,1]}(u),$$
$$L_2(u) = 30u^2(1 - u)^2(8 - 14u)I_{[0,1]}(u),$$
$$L_3(u) = 140u^3(1 - u)^3(10 - 18u)I_{[0,1]}(u).$$

All of the above kernels have $|E_C| < 0.3\%$. It appears that each of the kernels L_1, L_2, and L_3 produces quite wiggly OSCV curves for random samples generated from f^* and the $N(0,1)$ density.

The Figure 6 shows variety of robust and almost robust one-sided kernels of different shapes, but none of them performs satisfactory in the case of f^*. Thus, finding a kernel that improves practical performance of the method in the nonsmooth case appears to be an open

challenging problem.

5 Summary

The OSCV method for smooth density functions is proposed by Martínez-Miranda et al. (2009). In this article we extend the OSCV methodology to the case of nonsmooth densities. We also introduce the fully robust OSCV version that produces consistent bandwidths in both smooth and nonsmooth cases.

The proposed OSCV modifications, essentially, use different one-sided kernels to select the bandwidths by the cross-validation method. The selected bandwidths are then rescaled and plugged-in to the Gaussian density estimator. The nonsmooth density f^* plotted in Figure 1 **(a)** is used for discrimination of the proposed OSCV extensions.

We found and investigated many robust and almost robust kernel candidates for the fully robust OSCV implementation. Some of them are shown in Figure 6. None of these kernels performs satisfactory in the case of f^*. Moreover, the nonsmooth version of OSCV based on the one-sided Gaussian kernel L_G performs worse that the ordinary LSCV method in the case of f^*. Thus, practical implementations of the proposed theoretical extensions remain open to further research efforts.

The main problems experienced by the majority of the considered robust and almost robust one-sided kernels were selecting too variable bandwidths and/or producing multiple local minima in the OSCV curves. Similar difficulties were faced when implementing the ICV method (see Savchuk et al. (2010), Savchuk et al. (2011), and Savchuk (2017a)). This indicates that some nontraditional negative-valued cross-validation kernels may substantially improve the asymptotic properties of the cross-validation method while introducing challenging problems with their practical use.

The current implementation of the OSCV method in the smooth case of Martínez-Miranda et al. (2009) is based on the one-sided Epanechnikov kernel L_E. It appears that L_E produces inappropriately wiggly OSCV curves in the case of f^* (see Figure 2 **(b)**). Moreover, L_E frequently yields insufficiently smooth OSCV curves even in the case of the standard normal density (see Figure 4 **(a)**). The problem of L_E occasionally producing rough criterion curves persists for real data sets (see Figure 5 **(a)**). On the

other hand, we empirically found that for variety of smooth and nonsmooth densities and different sample sizes, the one-sided Gaussian kernel L_G usually produces smooth OSCV curves with one local minimum (see Figures 2 **(a)**, 4 **(b)** and 5 **(b)** for illustration). This indicates that L_G might be, potentially, superior than L_E for practical implementation of the OSCV method in the smooth case. This matter, however, requires further investigation that is out of scope of this article that is mainly devoted to extending OSCV to the case of nonsmooth densities.

The almost robust one-sided kernel shown in Figure 6 **(c)** with $(\alpha, \sigma) = (4, 0.8)$ has a nonnegative two-sided counterpart H_I. It then follows from the conclusions of Savchuk et al. (2013) and Savchuk and Hart (2017) that this kennel might, potentially, lead to successful implementation of the fully robust OSCV method in the regression context.

This article and the recent publication of Savchuk and Hart (2017) are supported by the R package `OSCV` that can be used for reproducing the presented results and allows for further experimenting in attempts of improving the OSCV method's practical implementation for smooth and nonsmooth density and regression functions.

References

A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360, 1984.

S.-T. Chiu. The effect of discretization error on bandwidth selection for kernel density estimation. *Biometrika*, 78(2):436–441, 1991.

D. B. H. Cline and J. D. Hart. Kernel estimation of densies with discontinuities or discontinuous derivatives. *Statistics*, 22(1):69–84, 1991.

W. Härdle. *Smoothing techniques*. Springer Series in Statistics. Springer-Verlag, New York, 1991. With implementation in S.

J. D. Hart and S. Yi. One-sided cross-validation. *Journal of the American Statistical Association*, 93(442):620–631, 1998.

M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection

for density estimation. *Journal of the American Statistical Association*, 91(433):401–407, 1996.

C. R. Loader. Bandwidth selection: classical or plug-in? *Annals of Statistics*, 27(2):415–438, 1999.

E. Mammen, M. D. Martínez Miranda, J. P. Nielsen, and S. Sperlich. Do-validation for kernel density estimation. *J. Amer. Statist. Assoc.*, 106(494):651–660, 2011. ISSN 0162-1459. doi: 10.1198/jasa.2011.tm08687. URL `http://dx.doi.org/10.1198/jasa.2011.tm08687`.

E. Mammen, M. D. Martínez Miranda, J. P. Nielsen, and S. Sperlich. Further theoretical and practical insight to the do-validated bandwidth selector. *J. Korean Statist. Soc.*, 43 (3):355–365, 2014.

M. D. Martínez-Miranda, J. P. Nielsen, and S. Sperlich. One sided cross validation for density estimation. In G. N. Gregoriou, editor, *Operational Risk Towards Basel III: Best Practices and Issues in Modeling, Management and Regulation*, pages 177–196. John Wiley & Sons, Hoboken, New Jersey, 2009.

M. Rudemo. Empirical choice of histograms and kernel density estimators. *Scandanavian Journal of Statistics*, 9(2):65–78, 1982.

O. Savchuk. *ICV: Indirect Cross-Validation*, 2017a. R CRAN package version 1.0.

O. Savchuk. *OSCV: One-Sided Cross-Validation*, 2017b. R CRAN package version 1.0.

O. Savchuk, J. Hart, and S. Sheather. An empirical study of indirect cross-validation. In *Nonparametric statistics and mixture models*, pages 288–308. World Sci. Publ., Hackensack, NJ, 2011.

O. Y. Savchuk and J. D. Hart. Fully robust one-sided cross-validation for regression functions. *Computational Statistics*, 2017. doi: 10.1007/s00180-017-0713-7.

O. Y. Savchuk, J. D. Hart, and S. J. Sheather. Indirect cross-validation for density estimation. *Journal of the American Statistical Association*, 105(489):415–423, 2010.

O. Y. Savchuk, J. D. Hart, and S. J. Sheather. One-sided cross-validation for nonsmooth regression functions. *J. Nonparametr. Stat.*, 25(4):889–904, 2013.

O. Y. Savchuk, J. D. Hart, and S. J. Sheather. Corrigendum to "One-sided cross-validation for nonsmooth regression functions" [J. Nonparametr. Stat., 25(4): 889-904, 2013]. *Journal of Nonparametric Statistics*, 28(4):875–877, 2016.

S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B*, 53(3):683–690, 1991.

B. W. Silverman. *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

B. van Es. Asymptotics for least squares cross-validation bandwidths in nonsmooth cases. *Annals of Statistics*, 20(3):1647–1657, 1992.

M. P. Wand and M. C. Jones. *Kernel smoothing*, volume 60 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London, 1995.

S. Yi. *On One-Sided Cross-Validation in Nonparametric Regression*. Ph.D dissertation, Texas A&M University, 1996.

6   Appendix

Notation.

For an arbitrary function g, define the following functionals:

$$\mu_2(g) = \int_{-\infty}^{\infty} u^2 g(u)\, du,$$

$$R(g) = \int_{-\infty}^{\infty} g^2(u)\, du,$$

$$D_g(z) = \int_{-\infty}^{z} g(u)\, du,$$

$$G_g(z) = \int_{-\infty}^{z} u g(u)\, du, \qquad z \in \mathbb{R}.$$

Based on D_g and G_g we define

$$B(g) = \int_{0}^{\infty} \left\{ z\big(1 - D_g(z)\big) + G_g(z) \right\}^2 dz + \int_{0}^{\infty} \left\{ z D_g(-z) + G_g(-z) \right\}^2 dz. \qquad (7)$$