QUANTITATIVE METHODS IN ECONOMICS
Vol. XV, No. 2, 2014, pp. 37 – 46

# COMPUTER-ASSISTED CHOICE OF SMOOTHING PARAMETER IN KERNEL METHODS APPLIED IN ECONOMIC ANALYSES

**Aleksandra Baszczyńska**
Department of Statistical Methods
University of Lodz
e-mail: albasz@uni.lodz.pl

**Abstract:** In the kernel method, it is necessary to determine the value of the smoothing parameter. Not without significance is the fact of using the objectivity in the selection of this parameter and a certain automation of the selection procedure, which is important especially for novice users of kernel methods in the process of statistical inference. In the paper some methods of choice of the smoothing parameter are presented with the results of the simulation study that indicate these methods of selecting the smoothing parameter as handy tool when kernel methods are used in economic analyses.

**Keywords**: kernel method, smoothing parameter, Silverman's practical rule, SiZer map

## INTRODUCTION

Kernel method is widely used in the estimation procedures of functional and numerical characteristics as well as in the hypothesis verification procedures concerning, for example, symmetry, goodness-of-fit or independence of random variables.

In all mentioned procedures it is necessary to determine the form of the weighting function known as the kernel function and the value of smoothing parameter which specifies the amount of smoothing in the kernel method. The parameters of kernel methods (kernel function and smoothing parameter) greatly influence the results of kernel methods.

In the case of kernel function basic attention is paid to the order of the kernel function which is closely connected with the number of vanishing moments and the number of existing derivatives of unknown density function [cf. Horová et al. 2012].

Classical kernel function, that is second order function, is unimodal and symmetric around zero density function [Gajek, Kałuszka 1996], [Wand, Jones 1995], [Silverman 1996]. It satisfies the following conditions [cf. Domański et al. 2014]:

$$\int_{-\infty}^{+\infty} K(u)du = 1, \tag{1}$$

$$\int_{-\infty}^{+\infty} uK(u)du = 0, \tag{2}$$

$$\int_{-\infty}^{+\infty} u^2 K(u)du = \mu_2(K) > 0. \tag{3}$$

Gaussian kernel function:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{u^2}{2}\right)}, \tag{4}$$

which is the kernel function with the conditions (1)-(3) is mostly used in practical applications.

Kernel function of $k$-th order, where $k$ is even number, fulfils the following:

$$\int_{-\infty}^{+\infty} K(u)du = 1, \tag{5}$$

$$\int_{-\infty}^{+\infty} u^l K(u)du = \mu_l(K) = 0 \quad \text{for } l = 1,...,k-1, \tag{6}$$

$$\int_{-\infty}^{+\infty} u^k K(u)du = \mu_k(K) \neq 0. \tag{7}$$

A large variety of kernel functions are presented in literature [e.g. Domański, Pruska 2000].

For $k = 2$ and $K(u) \geq 0$ kernel function is a density function. The kernel estimator of density function with such kernel function is also the density. For $k > 2$ kernel function may be negative what can cause that the density estimator may be negative.

Let $X_1, X_2, ..., X_n$ be continuous random variables with density function $f$. In parametric approach we assume that observations belong to one of the known density and the procedure of estimation means only estimation of parameters. In nonparametric approach we can use e.g. histogram or kernel Rosenblatt-Parzen estimator:

$$\hat{f}(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right), \tag{8}$$

where $K(u)$ denotes kernel function, $h$ is a smoothing parameter (bandwidth). Smoothing parameter, depending on the number of the sample $n$, $h=h(n)$: $\{h(n)\}$, is a sequence of a non-random positive numbers.

## CHOICE OF THE SMOOTHING PARAMETER

Choice of the smoothing parameter is a crucial issue in the kernel inference procedures. The subjective method is the simplest but time consuming one. In the case of density estimation, various kernel density estimators with different values of smoothing parameters are constructed and that parameter is chosen as the proper one which is used to construct "the best" estimator. This method is simply the judging by user's eye. Sometimes this method is treated as the pilot method of choosing the smoothing parameter in more complicated methods of selection of parameter. In the practical applications, not without significance is the fact of existing the objectivity in the selection of the smoothing parameter and a certain automation of the selection procedure, which is important especially for novice users of kernel methods in the process of statistical inference.

Silverman's rule of thumb is one of the mostly used objective method of smoothing parameter choice. When asymptotic mean integrated square error is used as the measure of closeness of the estimator $\hat{f}$ to the true density $f$, smoothing parameter which minimize this measure with the assumption that the unknown density is normal $N\left(0,\sigma^2\right)$ is the following:

$$\hat{h}_{SRT}=1.06\,\sigma n^{-\frac{1}{5}}, \tag{9}$$

where $\sigma$ can be estimated by $\hat{\sigma}=\sqrt{\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}}$ or $\hat{\sigma}_{IQR}=\frac{X_{\left\lceil\frac{3n}{4}\right\rceil}-X_{\left\lceil\frac{n}{4}\right\rceil}}{\Phi^{-1}\left(\frac{3}{4}\right)-\Phi^{-1}\left(\frac{1}{4}\right)}$,

for $\Phi^{-1}$ standard normal quantile function.

Terrell and Scott maximal smoothing method uses upper bound for the parameter calculated on the base of the asymptotic mean integrated square error. For specific value of scale, for example variance, this bound is attained by family of beta distributions $B(4,4)$ with variance $\sigma^2$ and it minimizes $\int_{-\infty}^{+\infty}\left(f^{(2)}(x)\right)^2dx$.

The unknown $\int_{-\infty}^{+\infty}\left(f^{(2)}(x)\right)^2 dx$ is estimated by its upper bound. The smoothing parameter is the following:

$$\hat{h}_{MSTS} = \frac{3}{35^5}\hat{\sigma}\left(\frac{R(K)}{\mu_2^2(K)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{10}$$

where: $R(K) = \int_{-\infty}^{+\infty} K(u)^2\,du$, $\mu_2(K) = \int_{-\infty}^{+\infty} u^2 K(u)du$.

Smoothing parameter selected by least squares cross-validation is the following:

$$\hat{h}_{LSCV} = \arg\min_{h\in H_n} LSCV(h), \tag{11}$$

where: $LSCV(h) = \int_{-\infty}^{+\infty} \hat{f}_h^2(x)dx - 2\frac{1}{n}\sum_{i=1}^{n} \hat{f}_{h,-i}(X_i)$ and

$\hat{f}_{h,-i}(X_i) = \frac{1}{(n-1)h}\sum_{j\neq i}^{n} K\left(\frac{X_i - X_j}{h}\right)$.

Smoothing parameter in the biased cross-validation method of selection has the form:

$$\hat{h}_{BCV} = \arg\min_{h} BCV(h), \tag{12}$$

where $BCV(h) = \frac{R(K)}{nh} + h^4\left(\frac{\mu_2(K)}{2}\right)^2 R\left(\tilde{\tilde{\tilde{f}}}^{(2)}\right)$ and

$R\left(\tilde{\tilde{\tilde{f}}}^{(2)}\right) = \frac{1}{n^2}\sum\sum_{i\neq j}(K'' * K'')(X_i - X_j)$.

In plug-in methods the pilot estimator with initial smoothing parameter is plugged in the formula of mean integrated square error. In this way the estimates of the unknown quantities are used. Smoothing parameter has the form:

$$\hat{h}_{DPI} = \left(\frac{R(k)}{\mu_2^2(K)\hat{R}\left(f^{(2)}\right)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}, \tag{13}$$

where $\hat{R}\left(f^{(2)}\right) == \frac{1}{n^2 g}\sum_{i,j=1}^{n} L^{(4)}\left(\frac{X_i - X_j}{g}\right)$ is the estimator with the kernel function $L$ and smoothing parameter $g$, where $L$ and $g$ may be different from $K$ and $h$.

In the iterative method the smoothing parameter is the following:

$$\hat{h}_{ITE} = \arg\min_{h} \hat{A}MISE(h), \tag{14}$$

where $\hat{A}MISE(h)$ is the suitable estimator of asymptotic mean integrated square error.

## SIZER MAP

SiZer map is another method of assessing the smoothing parameter, quite different from methods mentioned earlier. It presents not only one value of smoothing parameter used in kernel density estimation but the range of values. It is a graphical device which shows different structures of curves with different levels of smoothing. In this way it is possible to indicate significant features of estimator with respect to smoothing parameter used in the construction of density estimator.

SiZer map is used in the situation where we are interested in assessing the estimator of density function, especially which local maximums are true and which are false ones. It is a very good method to separate noise from the signal.

The family of kernel estimators for different values of smoothing parameter [cf. Baszczyńska 2014]:

$$\left\{\hat{f}_h(x): h \in [h_{\min}, h_{\max}]\right\} \tag{15}$$

are constructed, where $h_{\min} = 2B$, $B$ is the binwidth, $h_{\max} = x_{\max} - x_{\min}$.

All estimators in the family (15) are constructed with Gaussian kernel function for which number of points where derivative is zero decreases monotonically when the values of smoothing parameter is bigger.

A density estimator has derivatives equal to 0 at points of minimum, maximum and points of inflection. Before a point of minimum (maximum) the sign of the derivative $\frac{\partial^m E\left(\hat{f}_h(x)\right)}{\partial x^m}$ is positive (or negative), and after it, the derivative is negative (or positive). It is possible to identify structure of the estimator by zero crossings of the m*th* order of the derivative.

The hypothesis, verified in SiZer map, are the following [Chaudhuri, Marron 1999]:

$$H_0^{h,x} : \frac{\partial^m E\left(\hat{f}_h(x)\right)}{\partial x^m} = 0, \tag{16}$$

and

$$H_1^{h,x} : \frac{\partial^m E\left(\hat{f}_h(x)\right)}{\partial x^m} \neq 0. \tag{17}$$

The SiZer map is constructed in the following way: on the horizontal axis, values of $x$ are presented and on the vertical there are values of smoothing parameters. Different coloured regions are presented, where each colour has the special meaning. The following colour code is used: blue region means that $\hat{f}_h(x)$ is significantly increasing; red regions – $\hat{f}_h(x)$ is decreasing; purple regions – $\hat{f}_h(x)$ is not significantly increasing or decreasing; grey regions mean that the data are too sparse to make statements about the significance of increasing or decreasing of density estimator.

## SIMULATION STUDY

A simulation study was conducted to indicate some of the properties of the kernel density estimator, particularly the dependence of the density kernel estimation results on the parameters of the kernel method (kernel function and smoothing parameter) and, additionally, on the size of the sample. Seven sets of observations were regarded in the study:

1. random sample generated from population of normal distribution $N(0,1)$; $n=10$.
2. random sample generated from population of normal distribution $N(0,1)$; $n=30$.
3. random sample generated from population of normal distribution $N(0,1)$; $n=100$.
4. random sample generated from population with density of a mixture of two normal distributions $N(0,1)$ and $N(10,1)$ with equal weights; $n=10$.
5. random sample generated from population with density of a mixture of two normal distributions $N(0,1)$ and $N(10,1)$ with equal weights; $n=30$.
6. random sample generated from population with density of a mixture of two normal distributions $N(0,1)$ and $N(10,1)$ with equal weights; $n=100$.
7. 107 observations of the futures contract of price for units of carbon dioxide (CO2) in USD per tonne (current price from 2.01.2014 to 30.05.2014). Source of the data is the following: http://finanse.wp.pl/notowania-surowce-online.html.

In this way, different (small, medium and large) number of observations and in addition unimodal populations (sets (1)-(3)) and bimodal populations (sets (4)-(6)) were taken into account.

For all sets of observations the kernel density estimators were calculated with various kernel functions (e.g. classical kernels: Gaussian, Epanechnikov, uniform, quartic, triangular and kernel functions of higher order) and smoothing

parameter chosen in various way. The methods of choosing the smoothing parameter were the following: Silverman's rule of thumb SRT, Terrell and Scott maximal smoothing method MSTS, least squares cross-validation method LSCV, biased cross-validation BCV, direct plug-in method DPI, iterated method ITE.

The implementation of regarded methods was done using software with functions and toolboxes in Matlab (e.g. http://math.muni.cz/english/science-and-research/developed-software/232-matlab-toolbox.html; http://www.unc.edu/~marron/marron_software.html).

Values of smoothing parameters in kernel density estimation, calculated for all regarded sets of observations with Gaussian kernel function are presented in Table 1.

Table 1. Values of smoothing parameter in kernel density estimation with Gaussian kernel function

| Observations | Method of smoothing parameter selection | | | | | |
|---|---|---|---|---|---|---|
| | SRT | MSTS | LSCV | BCV | DPI | ITE |
| n=10 N(0,1) | 0.72008 | 0.77764 | 1.0504 | 1.5487 | 0.78788 | 1.1764 |
| n=30 N(0,1) | 0.60328 | 0.65151 | 0.7187 | 0.90829 | 0.72125 | 0.74501 |
| n=100 N(0,1) | 0.35267 | 0.38086 | 0.44758 | 0.4378 | 0.47666 | 0.41665 |
| n=10 N(0,1) and (10,1) | 3.6576 | 3.9500 | 1.4854 | 7.8638 | 2.8882 | 6.3832 |
| n=30 N(0,1) and (10,1) | 2.6542 | 2.8664 | 0.69337 | 5.7036 | 2.0929 | 0.87573 |
| n=100 N(0,1) and (10,1) | 2.0755 | 2.2414 | 0.50585 | 4.459 | 1.6562 | 0.54419 |
| 2.01.14-30.05.14 $CO_2$ price | 0.21279 | 0.2298 | 0.026502 | 0.21587 | 0.26041 | 0.18841 |

Source: own calculations

Comparing results of the kernel density estimations for samples from unimodal and bimodal populations gave the conclusion that in this second case the smoothing parameters are bigger. While the parameter controls the amount of smoothing, the issue of global or local smoothing parameter should be regarded. When a bandwidth is local, the amount of smoothing varies at each location. Around the modes the smoothing parameter should be smaller while in the tail of a distribution one can use much smoothing.

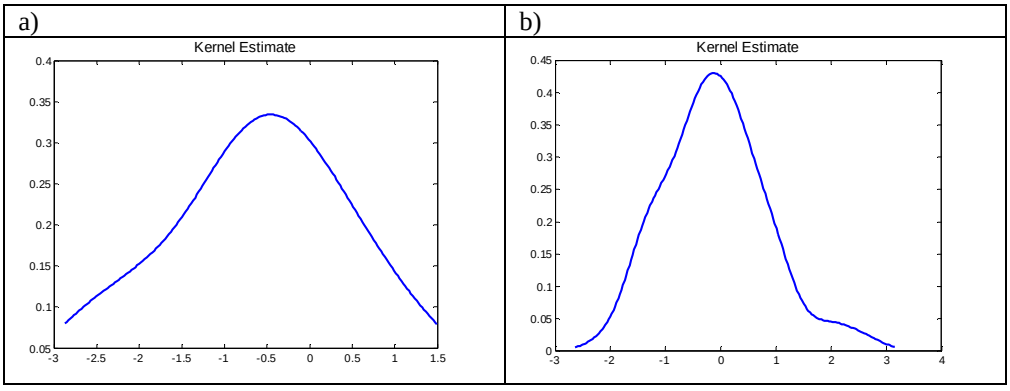
When biased cross-validation is used, the smoothing parameters for all sets of observations are bigger than in other methods of smoothing parameter selection.

It can be noticed that in most cases, the bigger the sample is, the bigger parameter is used in density estimation.

For the data of $CO_2$ price, the values are rather similar with exception of LSCV method (assuming independence of the observations). Similar results were obtained for other kernel functions.

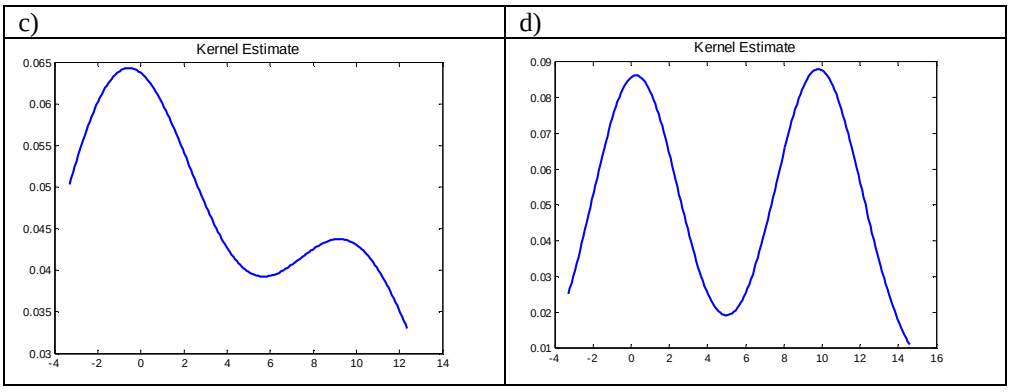
To deepen the analysis of kernel density estimation, including the significance of the sample size in the procedure of estimation, for all sets of observations kernel density estimators were calculated. Figures 1-3 present kernel density estimator where Gaussian kernel function and Silverman's rule of thumb were used for small and big sample sizes and for observations of the futures contract of price for units of carbon dioxide.

Figure 1. Kernel density estimator (Gaussian kernel function, Silverman's rule of thumb), data from unimodal populations a) n=10; b) n=100



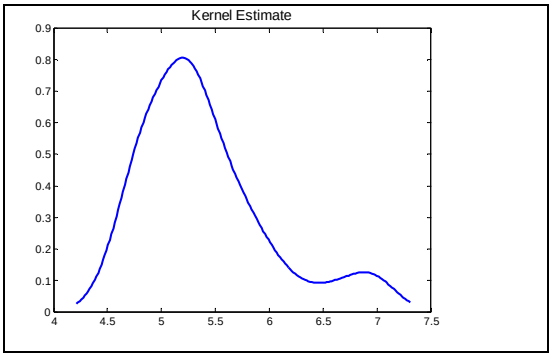

Source: own calculations

Figure 2. Kernel density estimator (Gaussian kernel function, Silverman's rule of thumb), data from bimodal populations c) n=10; d) n=100




Source: own calculations

Figure 3. Kernel density estimator (Gaussian kernel function, Silverman's rule of thumb), data of the futures contract of price for units of carbon dioxide

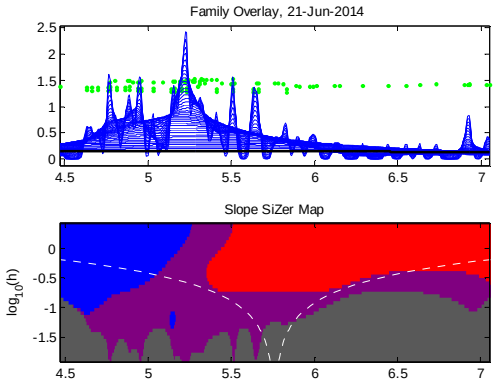



Source: own calculations

In the case of unimodal distribution, kernel density estimator is also unimodal, even for small sample. But bigger value of smoothing parameter should be used in this process of density estimation. The smaller size of the sample, the bigger smoothing parameter is used (cf. Table 1). When sample is chosen from bimodal distribution, only big sample ensures clearly bimodal density estimator, with smaller value of smoothing parameter.

Kernel density estimator for data of the futures contract of price for units of carbon dioxide indicates asymmetry and unimodality of the distribution. Additionally, for this set of observation the Sizer map was used. Figure 4 presents this SiZer map.

Figure 4. SiZer map, data of the futures contract of price for units of carbon dioxide




Source: own calculations

In SiZer map wide range of values of smoothing parameter is considered. For small values of smoothing parameter it is impossible to determine which maximums are true and which are false (for values up to approximately 0.0316, $\log_{10}(h) = -1.5$, grey region is observed). It means that the kernel density estimator is undersmoothed. For values of smoothing parameter bigger than 0.1 it is easy to inference that kernel estimator is positive or negative (blue regions and red regions). This result is consistent with previous result about values of smoothing parameter (cf. Table 1). The bigger value of smoothing parameter, the more smooth the estimator is.

## SUMMARY

Value of smoothing parameter in kernel density estimator is an issue of great importance. For users of kernel methods, it is very important that they can use some objective and automatic way of finding "the best" value of this parameter in practical applications. This condition is fulfilled by methods mentioned above with the help of specialised software. The presented results encourage not only scientists but also unexperienced users to apply them. But deeper analysis of "optimal" choice of smoothing parameter and kernel function is necessary in future research.

## REFERENCES

Baszczyńska A. (2014) Testing Significance of Peaks in Kernel Density Estimator by SiZer Map, Proceedings of the 8th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena, Foundation of the Cracow University of Economics, Kraków.

Chaudhuri P., Marron J.S. (1999) SiZer for Exploration of Structure in Curves, Journal of the American Statistical Association, 94, 447, 807 – 823.

Domański C., Pekasiewicz D., Baszczyńska A., Witaszczyk A. (2014) Testy statystyczne w procesie podejmowania decyzji, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Domański C., Pruska K. (2000) Nieklasyczne metody statystyczne, PWE, Warszawa.

Gajek L., Kałuszka M. (1996) Wnioskowanie statystyczne. Modele i metody, Wydawnictwa Naukowo-Techniczne. Warszawa.

Horová I., Koláček J., Zelinka J. (2012) Kernel Smoothing in MATLAB. Theory and Practice of Kernel Smoothing, World Scientific, New Jersey.

Silverman B.W. (1996) Density Estimation for Statistics and Data Analysis, Chapman & Hall, London.

Wand M. P., Jones M.C. (1995) Kernel Smoothing, Chapman & Hall, London.