SMOOTHING LEVEL SELECTION FOR DENSITY ESTIMATORS BASED ON THE MOMENTS

Rosa María García-Fernández[a] Department of Quantitative Methods for Economics and Business. University of Granada
ORCID 0000-0002-0642-5216
Federico Palacios-González . University of Granada

[a]Rosa María García-Fernández (corresponding autor)

Department of Quantitative Methods for Economics and Business

University of Granada

Faculty of Economics and Business Sciences, office C-112

Campus Universitario de Cartuja, 18071 Granada. Spain.

Phone: 34 958248789.

E-mail: rosamg@ugr.es

ABSTRACT

This paper introduces an approach to select the bandwidth or smoothing parameter in multiresolution (MR) density estimation and nonparametric density estimation. It is based on the evolution of the second, third and fourth central moments and the shape of the estimated densities for different bandwidths and resolution levels.

The proposed method has been applied to density estimation by means of multiresolution densities as well as kernel density estimation (MRDE and KDE respectively). The results of the simulations and the empirical application demonstrate that the level of resolution resulting from the moments method performs better with multimodal densities than the Bayesian Information Criterion (BIC) for multiresolution densities estimation and the plug-in for kernel densities estimation.

KEYWORDS: Multiresolution density estimation, kernel density estimation, bandwidth and level of resolution

1.INTRODUCTION

This paper develops a novel and straightforward approach to select the bandwidth or smoothing parameter in multiresolution models[1] estimation and nonparametric density estimation. It is based on the moments and the shape of the estimated densities for different resolution levels and bandwidths. The method is applied to density estimation by means of multiresolution densities [MRDE; see 12, 13 14] as well as kernel functions [KDE; see for instance, 18, 5].

Our choice of considering the moments for the selection of the smoothing parameter was guided by the observed changes in the shape of the estimated MRDE when the level of resolution, j , varies. If the level of resolution is too low, the density is smooth but the bias is large. Conversely, a large j leads to a rougher density but a small bias. Similar results are observed[2] when a density is fitted by the Kernel method. Let us suppose a bandwidth equals to $h = 2^{-j}$ with $j \in Z$. If j is too small then h is too large and the result

---

[1] The multiresolution densities are a particular case of semiparametric models (see, [12,14]).

[2] This is a well-known fact underlying all the bandwidth selection methods.

is a smooth and biased density. As j increases h decreases and the bias tends to diminish, but from a determined value of j, or its corresponding h, an undesirable roughness appears. To solve the problem, we need to select a value of j, or its corresponding smoothing parameter h, so that the bias will be reasonably small without incurring an excessive roughness.

In both methodologies, MRDE and KDE, the bias of the estimated density is shown clearly by its dispersion and shape, especially in the flattening of the fitted density. Or equivalently in an underestimation of the kurtosis that evolves towards more reasonable values as the resolution level j increases. That is, the bias evolution is related to the central moments of order 2, 3 and 4 since they are used to compute dispersion, asymmetry and kurtosis. When the resolution level in the MRDE increases, or the smoothing parameter in the KDE decreases, the flattening and the dispersion tends to stabilize indicating that the bias is small. From a certain level of resolution, the roughness begins to increase indicating which value of j or h should be selected. This value leads to a sufficiently smooth density with small bias. This is clearly shown in the graphs of sections 3, 4 and 5 that represent the evolution, as a function of j, of the expected value, variance, and skewness and kurtosis coefficients for MRDE and KDE.

The rest of the paper is organized as follows. Section 2 introduces the math expression to calculate the moments of a multiresolution density. Section 3 shows, by means of simulations, how to select the level of resolution using the moments of a MRDE. Section 4 extends the approach to the Kernel method. In both estimation methods, we use the Cubic Box Spline function defined in section 2.1. For the MRDE method, this is the scaling function generating the multiresolution analysis structure that contains the MRDE and their estimates. In the KDE method this function is used as the kernel. Section 5 contains an application to real data and section 6 concludes.

Finally, we want to point out that the MRDE is a technique devised for massive data. Therefore, everything that follows must be understood in a context of large sample sizes.

2. MOMENTS CALCULATION FOR A MRDE

This section shows the expressions of the central and non-central moments of a multiresolution density. The second, third and fourth central moments will be used to select the level of resolution of a MRDE. The math development is contained in the appendices A1-A6.

2.1. MULTIRESOLUTION DENSITIES

Let $\theta(x)$ be a symmetric density with mean zero and compact support $[-2,2]$, known as Cubic Box Spline. It is given by

$$\theta(x) = \begin{cases} 0 & if & x \leq -2 \\ p_1(2+x) & if & -2 \leq x \leq -1 \\ p_2(2+x) & if & -1 \leq x \leq 0 \\ p_2(2-x) & if & 0 \leq x \leq 1 \\ p_1(2-x) & if & 1 \leq x \leq 2 \\ 0 & if & x \geq 2 \end{cases},$$

where:

$$p_1(x) = \frac{x^3}{6} \text{ and where } p_2(x) = \frac{-x^3}{2} + 2x^2 - 2x + \frac{2}{3}. \qquad (1)$$

Applying dilations and translations to the density $\theta(x)$, the densities $\lambda_{j,k}(x)$ are built as follows [see 12]:

$$\lambda_{j,k}(x) = s\theta(sx - k), \qquad (2)$$

where $s = 2^j$. Note that $\lambda_{0,0}(x) \equiv \theta(x)$.

For each level of resolution j, the following MR densities:

$$f(x) = \sum_{k \in \mathbb{Z}} a_k \lambda_{j,k}(x), \qquad (3)$$

are defined.

In expression (3) $a_k \geq 0 \,\forall k \in \mathbb{Z}$ and $\sum_{k \in \mathbb{Z}} a_k = 1$. By definition, all these functions belong to the V_j space of the multiresolution analysis structure (MRA) defined by the scaling function $\theta(x)$ [see 7, 23, 11]. Any density of squared integrable, belonging to the space of Hilbert $L_2(\mathbb{R})$, has an approximation in each V_j.

In the definition of a multiresolution structure, dilations and translations of θ given by $\theta_{j,k}(x) = s^{1/2}\theta(sx - k)$ are used. It is obvious that $\lambda_{j,k}(x) = s^{1/2}\theta_{j,k}(x)$ and using this notation the MR density defined by (3) can be rewritten as follows:

$$f(x) = \sum_{k \in \mathbb{Z}} a_k \lambda_{j,k}(x) = s^{1/2} \sum_{k \in \mathbb{Z}} a_k \theta_{j,k}(x).$$

2.2 MOMENTS OF A RANDOM VARIABLE WITH MR DENSITY FUNCTION

The central moment of order r for a multiresolution (MR) density is calculated as follows:

$$\mu_f(r) = \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} \mu_{0,0}(i)\mu_a(r - i),$$

where $\mu_a(r-i) = \sum_{k\in\mathbb{Z}} a_k(k-\bar{k})^{r-i}$, $\bar{k} = \sum_{k\in\mathbb{Z}} k \times a_k$. By definition, the expression $\mu_{0,0}(r) = \mu_\theta(r)$ is the moment of order r of the density $\lambda_{0,0}(x) \equiv \theta(x)$ (see appendices A1 and A2). That is:

$$\mu_{0,0}(r) = \mu_\theta(r) = \int_{-2}^{2}(x-0)^r \theta(x)dx = \int_{-2}^{2} x^r \theta(x)dx \equiv m_\theta(r).$$

The calculus of $\mu_\theta(r)$ is in proposition 4 in appendix A1.

Given a MR density $f_j(x)$ the non-central moment of order r is defined by:

$$m_j(r) = \int_{-\infty}^{+\infty} x^r f_j(x)dx.$$

It can be obtained as follows (see appendix A4):

$$m_j(r) = \sum_{k\in Z} a_k \, m_{\lambda_{j,k}}(r),$$

where $m_{\lambda_{j,k}}(r) = \int_{-\infty}^{+\infty} x^r \lambda_{j,k}(x)dx = \frac{1}{s^r}\sum_{i=0}^{r}\binom{r}{i} m_\theta(i)k^{r-i}$.

2.3 ASYMPTOTIC PROPERTIES OF THE MOMENTS OF AN ESTIMATED MR DENSITY

PROPOSITION 1

Given a sample x_i $i = 1, 2, \ldots, n$ and an estimated MR density for that sample:

$$\hat{f}_j(x) = \sum_{k\in Z} \hat{a}_k \lambda_{j,k}(x),$$

for a value of j large enough is verified:

$$\hat{f}_j(x) = \frac{1}{n}\sum_{i=1}^{n} \lambda_{j,k(x_i)}(x),$$

where

$$k(x_i) = \begin{cases} \max\limits_{t\in Z \leq sx_i} t & if \ sx_i - \max\limits_{t\in Z \leq sx_i} t \leq 0.5 \\ \max\limits_{t\in Z \leq sx_i} t + 1 & if \ sx_i - \max\limits_{t\in Z \leq sx_i} t > 0.5 \end{cases}.$$

Proof. See Appendix A6.

PROPOSITION 2

Let be

$$\hat{m}_r(j) = \int_{-\infty}^{+\infty} x^r \hat{f}_j(x)\,dx,$$

the non-central moment of the estimated MR density from the sample x_i $i = 1, 2, \ldots, n$. If j converges to infinity, then $\widehat{m}_r(j)$ converges to its sample counterpart. That is:

$$\lim_{j \to \infty} \widehat{m}_r(j) = \frac{1}{n} \sum_{i=1}^{n} x_i^r \,.$$

Proof. See Appendix A6

Since each central moment of order r is a continuous (polynomic) function of the non-central moments of order less than or equal to r, we can state that as j approaches to infinity the central moments of a MR density estimated using a sample of size n also converge to the central moments of the sample.

3. MOMENTS METHOD FOR SELECTING THE RESOLUTION LEVEL TO ESTIMATE A MR DENSITY

In this section we introduce an alternative method to the Bayesian Information Criterion [17] to select the level of resolution in the estimation process of a MR density. It is based on the central moments of orders two, third, and four and the symmetry and kurtosis coefficients. When a MR density is estimated, if the resolution level is too low, the estimate is a smooth curve but it has excessive bias. Conversely, if the resolution level is too high the estimate bias is small but the roughness is large. In practice, the bias is mainly shown in an excessive dispersion and a flat density. Since the flattening can be measured by the Fisher coefficient, the evolution of the bias, as the resolution level increases, should be reflected in the gradual decrease of the central moment of order 2 and the central moment of order 4. Based on these moments, we will choose an appropriate resolution level so that the bias will be acceptable and the roughness of the estimator will be not excessive.

Since we are going to establish comparisons with the BIC criterion, let us introduce it briefly in the context of the MR densities. Any estimation using a MR density, for a finite size sample n and any resolution level j, can be considered a finite mixture of densities $\lambda_{j,k}(x)$ (see section 2.1) with the form:

$$\hat{f}_j(x) = \sum_{i=1}^{p} \hat{a}_{k_i} \lambda_{j,k_i}(x), \quad (4)$$

where $\hat{a}_{k_i}$ is the proportion of data within the interval $\left(\frac{k_i - 0.5}{2^j}, \frac{k_i + 0.5}{2^j} \right]$.

Note that the expression (4), estimator of (3), has a finite number of addends while (3) has infinite addends. This is explained as follows. Depending on the level of resolution, two extreme situations can arise. Firstly, j can be so small that the entire sample will be within a single interval $\left(\frac{k-0.5}{2^j}, \frac{k+0.5}{2^j}\right]$. In this case, there is only one coefficient $\hat{a}_k$ distinct from zero and mixture (4) degenerates into a single addend. Secondly, j can be so large that each observed value will be in a different interval existing as many $\hat{a}_{k_i}$ different from zero as different values are observed in the sample. That is, if n is the sample size and p is the number of different values observed in the sample, then the number of addends in the mixture (4) is p and it is verified that $1 \le p \le n$. Obviously, $p = n$ if all the sample values are different and each of them belongs to a single interval of the form[3] $\left(\frac{k_i-0.5}{2^j}, \frac{k_i+0.5}{2^j}\right]$ $k_i \in Z$.

Since mixture (4) contains p position parameters k_i $i = 1, 2, \dots, p$ and p mixture parameters $\hat{a}_{k_i}$ $i = 1, 2, \dots, p$ there are $m = 2 * p$ parameters. We can optimize the number of parameters, which depends on j, by using the BIC criterion [17]. That is, we will consider that the best value is that which minimizes the expression:

$$BIC(j) = -2Log(\hat{L}_j) + mLog(n),$$

where

$$\hat{L}_j = \prod_{i=1}^{n} \hat{f}_j(x_i),$$

is the sample likelihood of $\hat{f}_j(x)$.

The proposed method based on the moments, is simpler and requires less process time than the BIC. However, both criteria complement and reinforce each other, as will be shown in the following simulations.

To proceed, we have simulated a sample of size 10000 by using two generator models: a normal distribution and a mixture of double exponential distributions. We have fitted these generator models using MR densities for different levels of resolution in each case. Finally, we have calculated $E_j[X]$, $\mu_j(2)$, $\gamma_1 = \frac{\mu_j(3)}{\mu_j(2)^{\frac{3}{2}}}$ (Fisher coefficient of skewness or asymmetry) and $\gamma_2 = \frac{\mu_j(4)}{\mu_j(2)^2} - 3$ (Fisher coefficient of kurtosis) to select an

---

[3] Remind that these intervals form a partition of the real line and their amplitude converges to zero as j increases.

appropriate j. In the supplemental material, we provide more details about the calculus and a macro to apply the developed methodology to the data generating models used in this paper.

3.1. Normal distribution

Table 3.1 and Figure 3.1 shows the values $E_j[X], \mu_j(2), \gamma_1$ and γ_2 for a MR estimation of a N (10,5) distribution. Each value is divided by its empirical or sample counterparts, which are computed from the sample data without fitting any density. These indicators stabilize for $j = 0$ (see Figure 3.1). The BIC criterion provides the value $j = -1$. Figure 3.2 displays the fitted densities ($\hat{f}(x)$) for $j = -1$ and $j = 0$, and the data generator model ($N(10,5);\ f(x)$).

Table 3.1. Ratios between MRDE moments and sample moments

j	E[X]	Variance	Asymmetry	Kurtosis
-3	1.0014	2.0494	0.5553	1.987432
-2	1.0058	1.2919	-0.6275	-3.905073
-1	1.0011	1.0717	0.0905	-2.285609
0	1.0000	1.0160	1.0618	1.212460
1	0.9998	1.0033	1.0149	0.887280
2	1.0000	1.0010	1.0179	1.072510
3	1.0000	1.0006	0.9854	1.023095
4	1.0000	1.0001	0.9973	0.985362

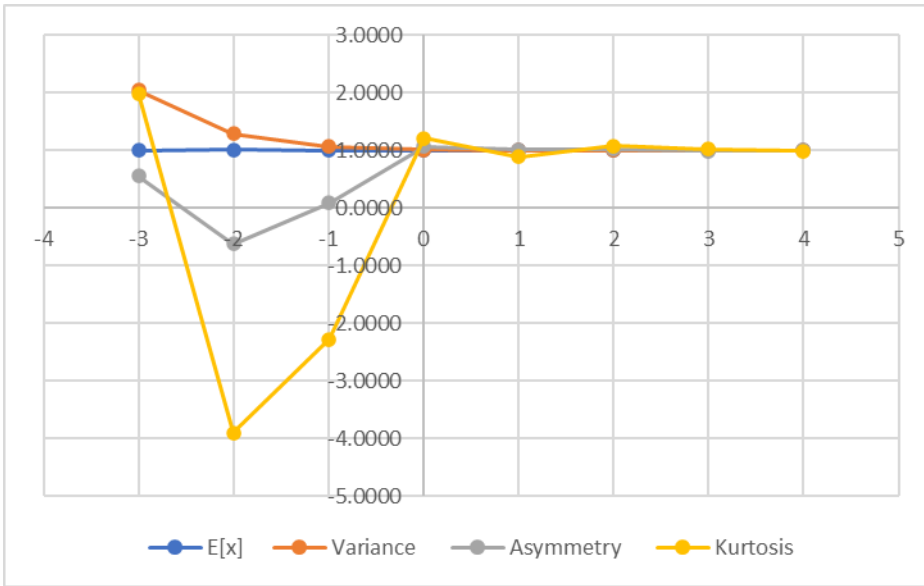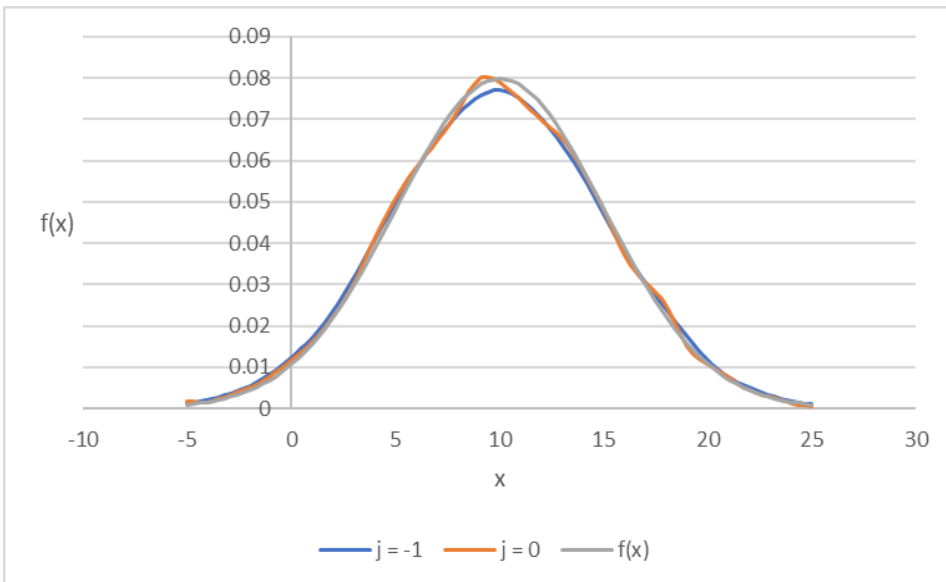Fig. 3.1. Ratios between MRDE moments and sample moments

Fig. 3.2 Estimated densities $\hat{f}(x)$ for $j = -1, j = 0$ and data generator model $f(x)$

The estimate for $j = -1$ is smooth but the bias is noticeable. For $j = 0$ the bias almost disappears but there is a small roughness that may be acceptable. According to the BIC criterion, the optimum level of resolution is $j = -1$.

3.2. Mixture of double exponential distribution

The density function of a double exponential distribution with parameters μ and θ is given by:

$$f(x) = e^{-\frac{|x-\mu|}{\theta}} \quad \forall x \in \mathbb{R}.$$

In this illustration, the generator model will be a mixture of three densities of this type whose parameters are in Table 3.2.

Table 3.2. Parameters of the double exponential distribution

μ	θ	π
20	5	0.3
30	6	0.5
40	7	0.2

Table 3.3 and Figure 3.3 display the values of $E_j[X]$, $\mu_j(2)$, γ_1, γ_2, divided by their empirical counterparts, using levels of resolution from $j = -3$ to $j = 4$. As can be seen, stability is reached either when $j = -1$ or $j = 0$. Figure 3.4 shows the estimations, $\hat{f}(x)$) for $j = -2$, $j = -1$, $j = 0$, and the data generator model (mixture of double exponential, $f(x)$).

Table 3.3. Ratios between MRDE moments and sample moments

j	E[x]	Variance	Asymmetry	Kurtosis
-3	1.00020842	1.23072002	0.64137168	0.56500524
-2	1.0000844	1.05642867	0.92675066	0.90709002
-1	1.00058047	1.0155786	0.97057506	0.97393645
0	0.99986737	1.00355414	0.99758049	0.9933315
1	0.99996728	1.000518	0.99559745	0.99654088
2	0.99999484	1.00019494	1.00076368	1.00119058
3	1.00000043	1.00001854	0.99983838	1.00028379
4	1.00000388	1.00003893	0.99975579	0.99954511

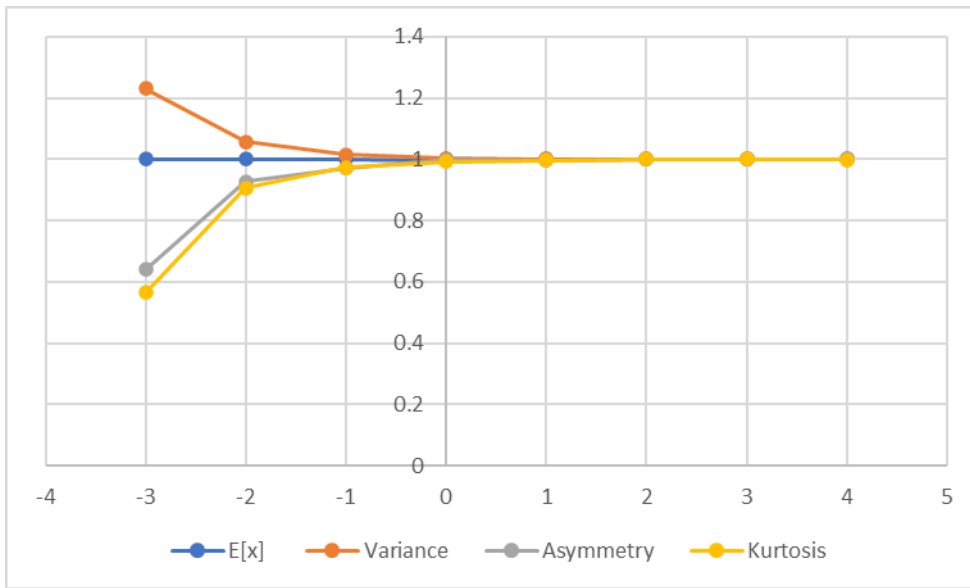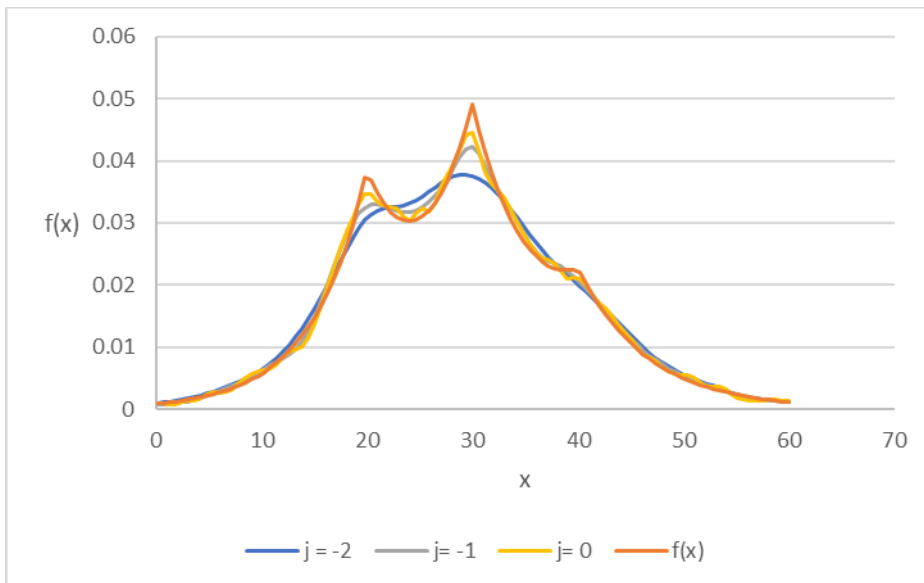Fig. 3.3 Ratios between MRDE moments and sample moments

Fig. 3.4. Estimated densities $\hat{f}(x)$ for $j = -2, j = -1, j = 0$ and data generator model $f(x)$

The mixture has three modes that are difficult to capture by the estimations[4]. This leads us to use higher resolution levels and rougher estimates to avoid the bias that such difficulty produces. Based on Figure 3.4, we would opt for $j = -1$ discarding $j = -2$

---

[4] Unless this is done parametrically using the EM algorithm on a mixture model of three double exponential distributions. But for a sample of size 10000 the process time is too long.

because of bias excess, and $j = 0$ for roughness excess. In this example, certain difficulties are encountered in the BIC criterion. Due to the fact it is based on the principle of parsimony, it tends to give up the peaks and selects a smoother estimation. The level of resolution, according to the BIC is $j = -2$.

4. MOMENTS AND SELECTION OF THE BANDWIDTH OF A KERNEL DENSITY

Kernel density estimation (KDE) has become a common and useful tool for empirical studies. The discussion on the selection of the bandwidth has given rise to numerous publications on the subject. Nonetheless, part of the scientific community that works in nonparametric statistics has accepted that it may not be a perfect procedure for selecting the optimal bandwidth. We will not give an overview of kernel estimation techniques since our main aim is to extend the use of the moments to the choice of bandwidth parameter. We refer readers to [15, 2, 9, 6 ,24, 4, 10, 21] for a review of the bandwidth selector techniques.

4.1 MOMENTS OF A KERNEL DENSITY

The Kernel estimator of a density for a sample $x_i \ i = 1,2,\dots,n$ is given by:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} \theta\left(\frac{x - x_i}{h}\right),$$

where h is the bandwidth or smoothing parameter and θ is the cubic box spline that we are going to use as the Kernel function.

For this density, it is verified (see appendix A5):

$$E_{\hat{f}_h}[X] = \frac{1}{n} \sum_{i=1}^{n} x_i,$$

and

$$\mu_{\hat{f}_h}(r) = \sum_{t=0}^{r} \binom{r}{r} h^t m_\theta(t) \mu(r - t),$$

where $m_\theta(t)$ is the non-central moment of order t of the density θ (see section 2 and the appendix A5) and $\mu(r - t)$ is the central moment of order $r - t$ for the sample. That is:

$$\mu(r - t) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^{t-r}.$$

4.2 SELECTION OF THE BANDWIDTH OF A KERNEL DENSITY BASED ON THE MOMENTS

The Kernel estimator of a density for a sample x_i $i = 1,2, \dots, n$ [5] is:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right), \quad (5)$$

where K is the kernel and h is the so-called smoothing parameter.

Let us assume that $h = \frac{1}{2^j}$ where j is an integer number. The kernel function that we are going to use is θ, that is, the cubic box spline introduced in section 2.1. It is evident that (5) can be written as follows:

$$\hat{f}_h(x) = \frac{2^j}{n} \sum_{i=1}^{n} \theta\left(2^j x - 2^j x_i\right) = \frac{1}{n} \sum_{i=1}^{n} 2^j \theta\left(2^j x - 2^j x_i\right). \quad (6)$$

Utilizing the multiresolution analysis notation, the expression (6) can rewrite as:

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} \lambda_{j,2^j x_i}(x). \quad (7)$$

The estimator of a MRDE at the resolution level j defined in (4) is:

$$\hat{f}_j(x) = \sum_{l=1}^{p} \hat{a}_{k_l} \lambda_{j,k_l}(x), \quad (8)$$

where $\hat{a}_{k_l} = \frac{n_l}{n}$, n_l is the number of sample values x_l such that k_l is the closest integer to $2^j x_l$. This allows us to write (8) as:

$$\hat{f}_j(x) = \frac{1}{n} \sum_{l=1}^{p} n_l \lambda_{j,k_l}(x). \quad (9)$$

Assuming that $k(x_l)$ is the closest integer to $2^j x_l$ it is easy to understand that:

$$\hat{f}_j(x) = \frac{1}{n} \sum_{i=1}^{n} \lambda_{j,k(x_i)}(x). \quad (10)$$

Note that expressions (9) and (10) are equal. We can obtain (10) from (9) by a frequency count on the values $k(x_i), i = 1,2, \dots n$. In (9) p represents the number of values $k(x_i)$ found and $n_l, l = 1,2, \dots p$, is the number of repetitions observed for each of them. Therefore, we can write:
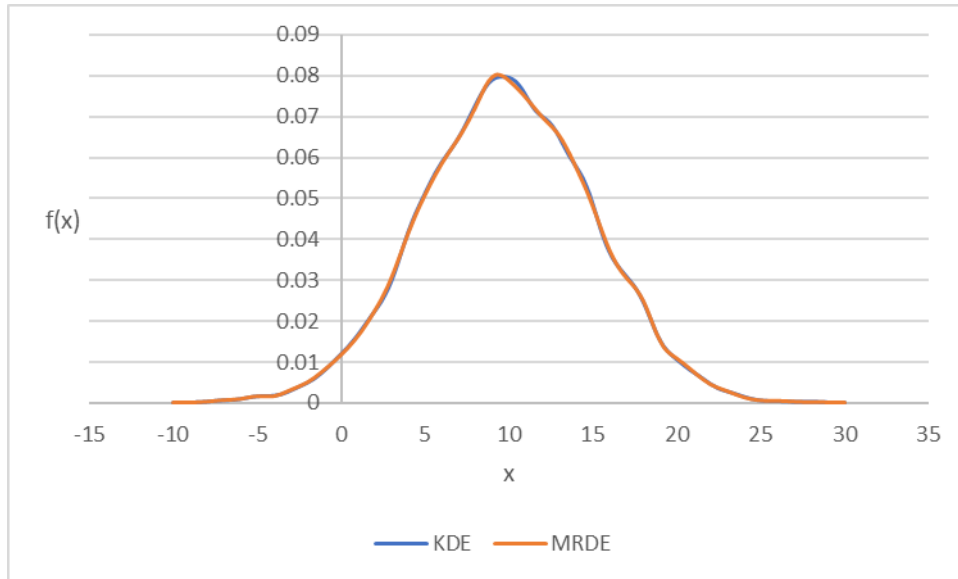
$$\hat{f}_j(x) = \frac{1}{n} \sum_{i=1}^{n} 2^j \theta\left(2^j x - k(x_i)\right),$$

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} 2^j \theta \left(2^j x - 2^j x_i \right),$$

Taking into account that by definition $\lfloor k(x_i) - 2^j x_i \rfloor \leq 0.5$ for a j sufficiently high both expressions must give very close results.

To illustrate the selection of h based on the moments, we have simulated a sample of size 10000 from a N (10, 5). Figure 4.1 shows the MRDE and the KDE for $j = 0$.

Fig. 4.1 MRDE and KDE for $j = 0$

As can be seen in the Figure 4.1, both estimates are very similar and show a similar degree of roughness. This fact suggests to us that the moment method is suitable to select an appropriate h when $h = 1/2^j$ with j integer. Table 4.1 and Figure 4.2 show $E_j[X]$, $\mu_j(2)$, γ_1 and γ_2 divided by their empirical counterparts for $j = -3, \ldots, 1, 2, 3, 5$. The moments have been calculated by the expression developed in this section using the fitted densities for the values of h that correspond to the above j. The values of j are on the abscissa axis.

Table 4.1 Ratios between KDE moments and sample moments

j	E[x]	Variance	Asymmetry	Kurtosis
-3	1	1.83770525	0.40140818	4.41595072
-2	1	1.20942631	0.75184944	1.27816406
-1	1	1.05235658	0.92630844	0.95204746

14

0	1	1.01308914	0.98068268	0.97763648
1	1	1.00327229	0.99511157	0.99369833
2	1	1.00081807	0.99877415	0.99837911
3	1	1.00020452	0.9996933	0.99959192
4	1	1.00005113	0.99992331	0.9998978
5	1	1.00001278	0.99998083	0.99997444

Fig. 4.2. Ratios between KDE moments and sample moments
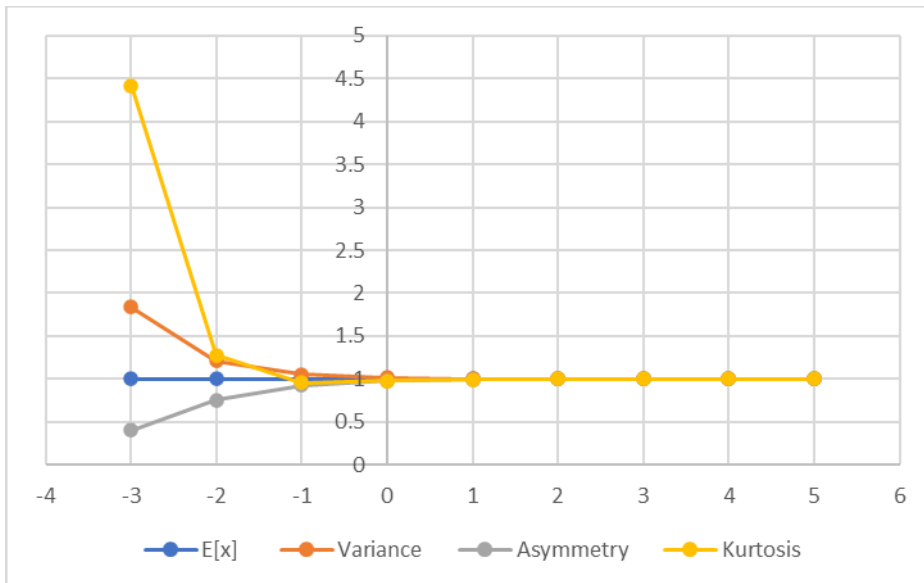
Note that the stability of both moments is reached when $j = 0$ which is equivalent to $h = 1$. Figure 4.3 displays the MR and the kernel estimations for $j = -1$ and $j = 0$.

Fig. 4.3 Estimated densities $\hat{f}(x)$ for $j = -1, j = 0$ and data generator model $f(x)$
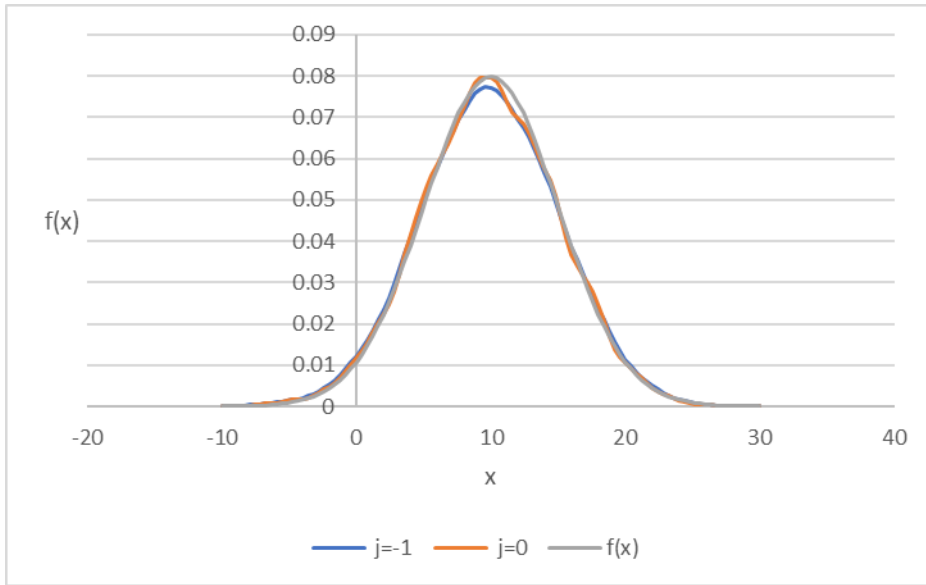
Note that j is conveniently the same for both estimates and it can be obtained from the MR or kernel estimated moments. KDE and MRDE estimates are similar since both are good approximations of the same unknown density.

In the following simulation, the generator model is a mixture of three double exponential distributions whose parameters are in Table 3.2. Table 4.2 and Figure 4.4 display the values of $E_j[X]$, $\mu_j(2)$, γ_1, γ_2, divided by their empirical counterparts, using levels of resolution from $j = -3$ to $j = 4$. As can be seen, stability is reached either when $j = -1$ or $j = 0$. Figure 4.5 shows the estimations, $\hat{f}(x)$) for $j = -1, j = 0$, and the data generator model (mixture of double exponential distributions, $f(x)$).

Table 4.2 Ratios between KDE moments and sample moments

j	E[x]	Variance	Asymmetry	Kurtosis
-3	1.00020842	1.23072002	0.64137168	0.56500524
-2	1.0000844	1.05642867	0.92675066	0.90709002
-1	1.00058047	1.0155786	0.97057506	0.97393645
0	0.99986737	1.00355414	0.99758049	0.9933315
1	0.99996728	1.000518	0.99559745	0.99654088
2	0.99999484	1.00019494	1.00076368	1.00119058
3	1.00000043	1.00001854	0.99983838	1.00028379
4	1.00000388	1.00003893	0.99975579	0.99954511

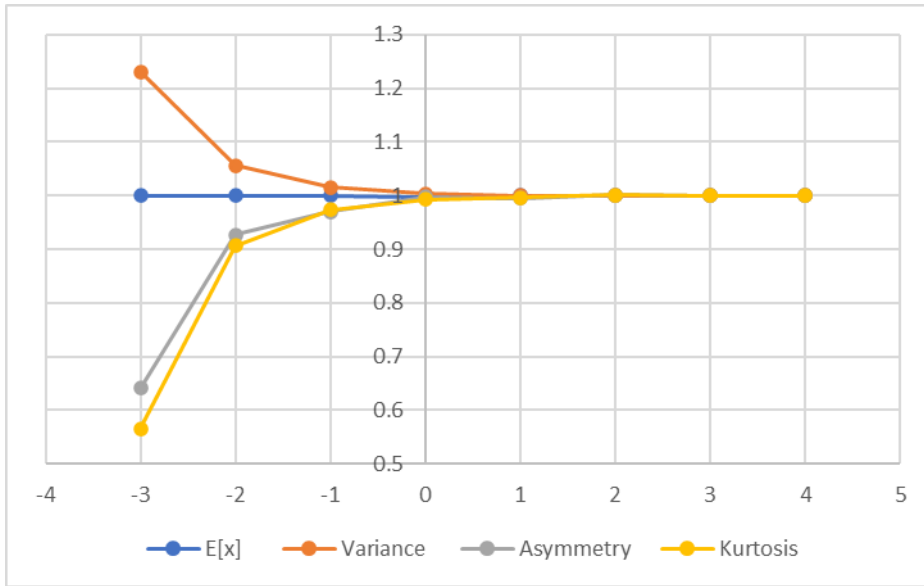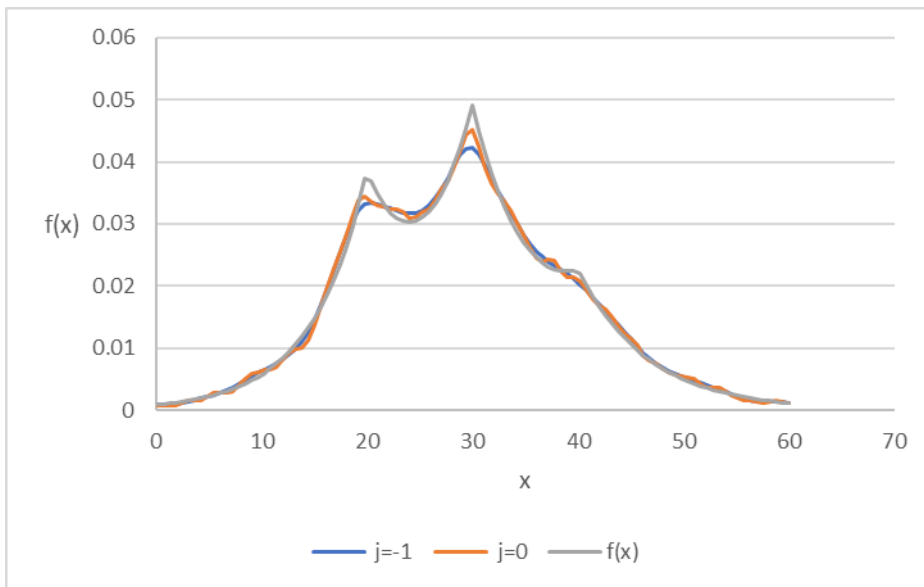Fig. 4.4. Ratios between KDE moments and sample moments

Fig. 4.5 Estimated densities $\hat{f}(x)$ for $j = -1, j = 0$ and data generator model $f(x)$

An alternative way to compare KDE and MRDE is making $2^j = 1/h$ in (10). That is:

$$\hat{f}_j(x) = \frac{1}{n}\sum_{i=1}^{n} \lambda_{j,k(x_i)}(x) = \frac{1}{n}\sum_{i=1}^{n} 2^j \theta\left(2^j x - k(x_i)\right) = \frac{1}{nh}\sum_{i=1}^{n} \theta\left(\frac{x}{h} - k(x_i)\right)$$

$$= \frac{1}{nh}\sum_{i=1}^{n} \theta\left(\frac{x - hk(x_i)}{h}\right). \quad (11)$$

Observe that (5) and (11) are quite similar. Taking into account that by definition:

$$k(x_i) - 0.5 < 2^j x_i \leq k(x_i) + 0.5,$$

and multiplying by $h = 2^{-j}$ the three terms of the above inequality we have:

$$hk(x_i) - h0.5 < x_i \leq hk(x_i) + h0.5.$$

The latter expression shows an increasing approximation between x_i and $hk(x_i)$. Note that the amplitude of the previous interval is h.

$$\lim_{h \to 0} hk(x_i) = \lim_{j \to \infty} \frac{k(x_i)}{2^j} = x_i.$$

The MRDE and the KDE had been compared in terms of time needed to run their density function in [12]. Nonetheless, the previous simulations reveal some facts that are worth highlighting. The MR density is not a particular case of kernel density. On the one hand, when the multiresolution densities are estimated according to (4), the results are similar to a modified kernel in which each sample data x_i is substituted in (5) by $hk(x_i)$ to obtain (6), with $h = 2^{-j}$. On the other hand, we cannot state that a kernel estimator is a multiresolution kernel estimator. We could make the kernel and the scaling function identical. Also, we can equal both dilation factors by making $h = 2^{-j}$. But the kernel for $h = 2^{-j}$ will be a density of the V_j space of the multiresolution analysis structure only if the sample is of the form $x_i = \frac{k_i}{2^j}$ $i = 1,2,..,n$, with $k_i \in Z \; \forall i = 1,2,...n$ and where j is a fixed integer determined by the sample. Any other estimation with a different h will no longer be a function of the multiresolution structure.

Despite the above comment, we have to point out that there is a well-developed theory about the generalized kernel estimators, developed from the wavelets and the multiresolution analysis structures (see for instance [22]). Broadly speaking, this methodology requires the mother wavelet or scaling function of the multiresolution analysis structure to generate orthogonal bases for the V_j spaces of the MRA. This is not the case of the cubic box spline since it generates non orthogonal Riesz Bases. Going deeper into this aspect is an interesting question, but it is out of the scope of this work. 5. REAL DATA APPLICATION

In this section, we apply the proposed method to the gross income of the Spanish households. The sample data comes from the Spanish Survey of Household Finances (EFF) for the year 2014, which was conducted by the Bank of Spain [1]. The EFF provides information on assets, debt, income and spending. The sample size is 6,120 households. The household income is calculated as the sum of labor and non-labor incomes for all household members in 2013. It is expressed in hundred thousand euros.

Table 5.1 and Figure 5.1 show the evolution, according to j, of $E_j[X]$, $\mu_j(2)$, γ_1, and γ_2, divided by their empirical counterpart for levels of resolution from $j = -13$ to $j = -6$.

Table 5.1 Ratios between MRDE moments and sample moments

j	E[X]	Variance	Asymmetry	Kurtosis
-13	0.99717924	1.01153181	0.98738403	0.98380212
-12	0.9995666	1.00258459	0.99586171	0.99430889
-11	0.99999127	1.00037759	0.9990044	0.99853096
-10	1.00016918	0.99997439	0.99958258	0.99926656
-9	0.99990519	1.00016536	0.99988219	0.99980877
-8	0.99989802	0.99998004	1.0000367	1.00010499
-7	0.99995971	0.99999305	1.00003152	1.0000789
-6	0.99992814	1.00000293	0.9999922	0.99998317

Fig. 5.1. Ratios between MRDE moments and sample moments
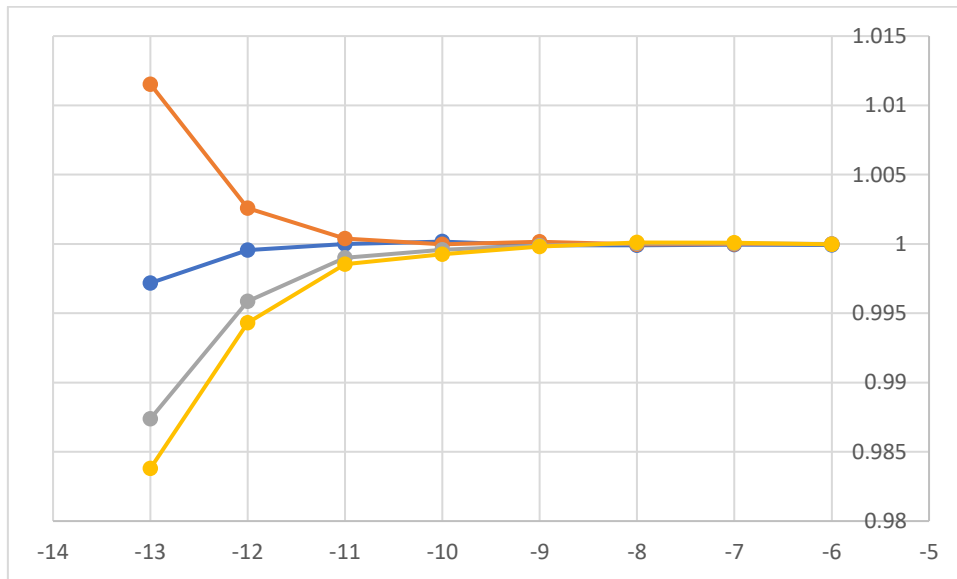
According to the BIC criterion the optimum is $j = -13$. However, the moments stabilize for $j = -9$ or $j = -8$ (Figure 5.1). Let us focus on this difference by comparing the density estimates for the two resolution levels plotted in Figure 5.2.

Fig. 5.2 Estimated multiresolution density   for $j = -8$ and $j = -13$
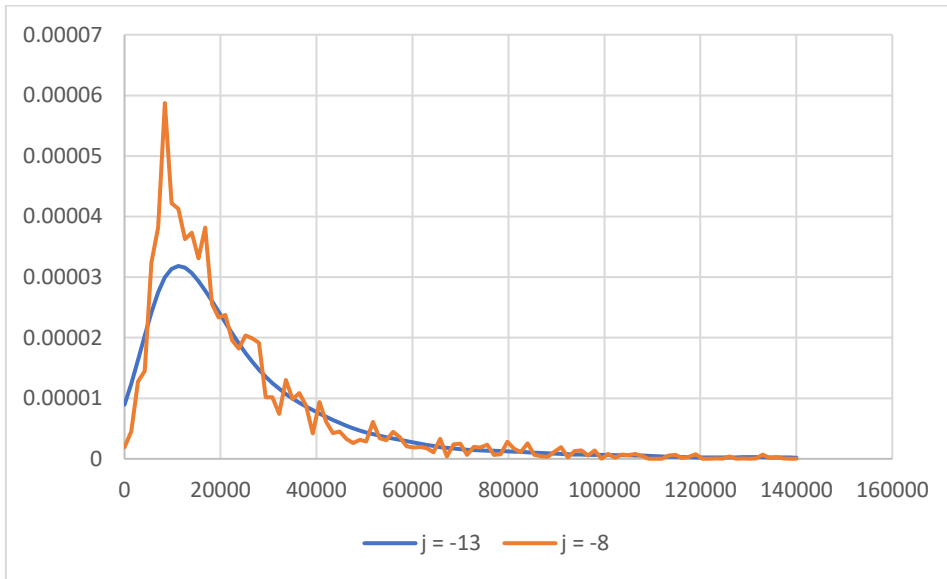
The density has a peak between 8000 and 9000 euros that cannot be captured accurately by using $j = -13$. So, a higher level of resolution, $j = -8$ or $j = -9$, is needed. The bias for $j = -13$ is evident when the two densities are compared. The roughness for $j = -8$ is clearly appreciable. The BIC chooses the smoothness of the curve, which leads to a very skewed estimated density around the mode. A similar fact has been shown in section 3.2. We have observed empirically that roughness has only a slight effect on the cumulative distribution function. Nonetheless, the bias has a remarkable impact on the concentration measurement producing an underestimation of the Gini index and the Lorenz curve. This is an important issue to be considered if we study distributional aspects of the distribution as concentration or inequality through the fitted density. At this point, it should be noted that the kernel method is frequently applied to study income distribution (see for instance [8,16,3,20]). Figures 5.3 and 5.4 plot the cumulative distribution functions and the Lorenz curves respectively. The cumulative distribution functions are similar except in the income interval [0, 10000]. This difference leads to an underestimation of the Gini index[5] : for $j = -13$ the index equals to 0.4338 and for $j = -8$ it is equals to 0.5131. It also affects the Lorenz curve (see figure 5.4) which is underestimated for $j = -13$. Therefore, the level of resolution $j = -8$

---

[5] Note that the values for the Gini coefficient can differ from other publications since our illustration is based on gross income instead of net income.

obtained by the method of moments is preferable to the value $j = -13$ selected by the BIC in the situations set out above.

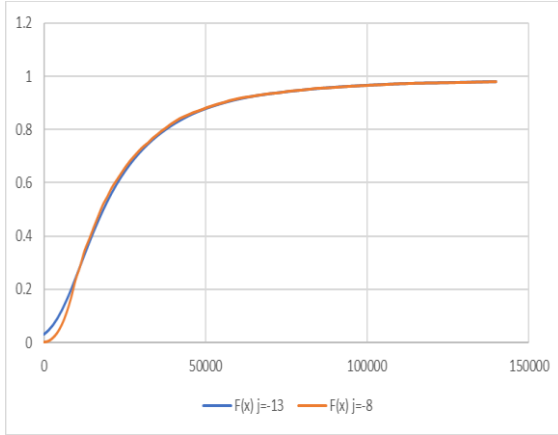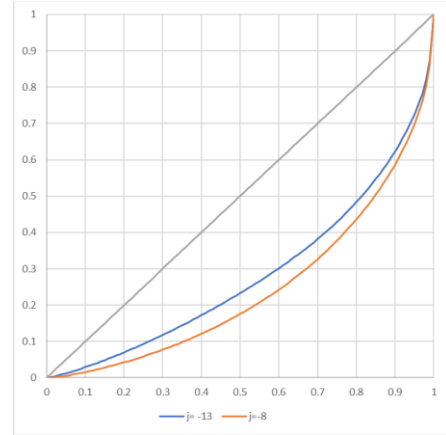Fig. 5.3. Cumulative distribution functions   Fig. 5.4. Lorenz curves


Next, we repeat the estimation for kernel densities and compare the results. Table 5.2 Figure 5.5 show the evolution, according to j, of $E_j[X]$, $\mu_j(2)$, γ_1 and γ_2 divided by their sample counterpart for levels of resolution from $j = -13$ to $j = -5$.

Table 5.2 Ratios between KDE moments and sample moments

j	$h = 1/2^j$	E[X]	Variance	Asymmetry	Kurtosis
-13	8192	1	1.00759984	0.98870759	0.98497182
-12	4096	1	1.00189996	0.99715681	0.99621088
-11	2048	1	1.00047499	0.99928794	0.9990507
-10	1024	1	1.00011875	0.99982191	0.99976255
-9	512	1	1.00002969	0.99995547	0.99994063
-8	256	1	1.00000742	0.99998887	0.99998516
-7	128	1	1.00000186	0.99999722	0.99999629
-6	64	1	1.00000046	0.9999993	0.99999907
-5	32	1	1.00000012	0.99999983	0.99999977

Fig. 5.5 Ratios between KDE moments and sample moments
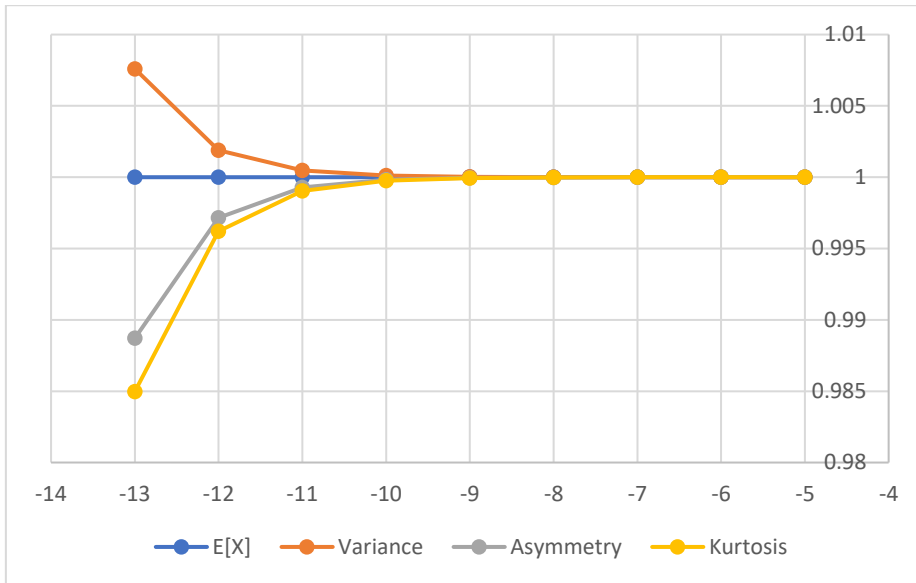
The appropriate level of resolution according to the moments is $j = -10$ or $j = -9$.

Fig. 5.6 Estimated kernel density for $j = -9$ and $j = -13$
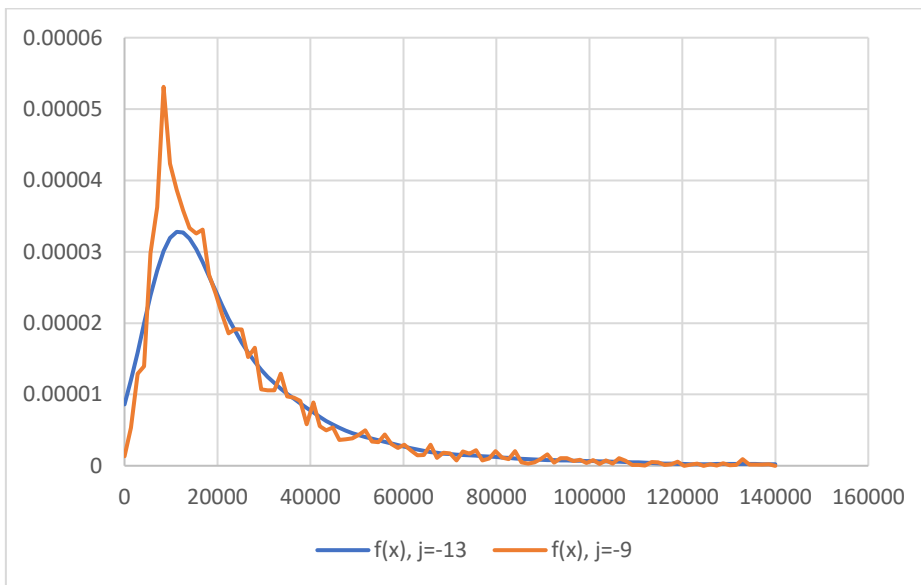
     The results are similar reinforcing the idea of applying the analysis performed on the MRDE. The level of resolution selected for the MRDE was $j = -9$. For the KDE we have opted for $h = \frac{1}{2^{-9}} = 2^9 = 512$. The plug-in method to select the optimum h [19] provides the result $h = 459.277$. This value corresponds to $j = -\log_2 459.277 = -8.8432$ which rounded to the nearest integer would give $j = -9$. If we use the MRDE for $j = -13$ the plug-in method provides the result $h = 3140.324$. That is $j = -\log_2 3140.324 = -11.6167$ whose nearest integer is $j = -12$. This is less

conservative than the BIC but still conservative. In any case, the resulting j is the same or it is very close to that used for the plugged MRDE.

Generalizing, if we use an estimated MR for a given j as plugged density, the plug-in method provides a value of h equal to $-\log_2 h$ whose nearest integer is the value of j utilized to estimate the MRDE. It is faster and easier to use the method of the moment for KDE and determine the $h = \frac{1}{2^j}$ that we will use in the estimation. If we want more conservative results, regarding the smoothness of the fit, we can reduce the value of j by one or two units, paying special attention to the increasing bias.

6. CONCLUSIONS

This paper introduces an approach to select the bandwidth or smoothing parameter in semiparametric and nonparametric density estimation. It is based on the evolution of the expected value, the variance, the symmetry and kurtosis coefficients of the estimated densities for different bandwidths. Using these values, divided by their empirical counterpart, we select a resolution level so that the bias will be acceptable and the roughness of the estimator will be not excessive.

This method has been applied to the density estimation by means of multiresolution densities as well as Kernel density estimation. In this way, we have expanded the available criteria to smoothing parameter selection.

The results of the simulations and the empirical application indicate that the level of resolution resulting from the moments method is more flexible to fit a multimodal distribution than those resulting from the BIC for MRDE and the plug-in for KDE. The BIC chooses the smoothness of the curve which leads to a skewed estimated density around the modes. The method of the moments attributes more importance to the use of higher resolution levels and hence rougher estimates to avoid the bias that the fitting produces. This procedure is recommended to analyse some distributional aspects such as the concentration of income. As it has been shown in the empirical application, the bias can produce an underestimation of the concentration of the distribution.

REFERENCES

[1] Banco de España. Survey of Household Finances (EFF) 2014: Methods, Results and Changes since 2011, Analytical article, 24, January 2007.

[2] R. Cao, A. Cuevas and W. Gonzalez Manteiga, *A comparative study of several smoothing methods in density estimation*, Comput. Stat. Data Anal. 17 (1994), pp. 153–176. https://doi.org/10.1016/0167-9473(92)00066-Z

[3] A. Charpentier and E. Flachaire, *Log-Transform Kernel Density Estimation of Income Distribution*.,L'Actualité économique, 91 (2015), pp. 141–159. https://doi.org/10.7202/1036917ar

[4] P. Hall and J.S. Marron, *Estimation of integrated squared density derivatives*, Stat. Probab. Lett. 6, (1987), pp. 109–115. https://doi.org/10.1016/0167-7152(87)90083-6

[5] W. Härdle, (1991). *Smoothing techniques*. Springer, New York, 1987.

[6] N.B.]Heidenreich, A. Schindler and S. Sperlich, *Bandwidth selection for kernel density estimation: a review of fully automatic selectors*, AStA Advances in Statistical Analysis 97 (2013), pp. 403–433. https://doi.org/10.1007/s10182-013-0216-y

[7] E. Hernández and G. Weiss, *A first course on wavelets*, CRC Press. New York, 1996. https://doi.org/10.1201/9780367802349

[8] S.P.Jenkins, *Did the middle class shrink during the 1980s? UK evidence from kernel density estimates,* Economics Letters 49 (1995), pp. 407-413. https://doi.org/10.1016/0165-1765(95)00698-F

[9] M.C. Jones, *On correcting for variance inflation in kernel density estimation*. Comput. Statist. Data Anal. 11 (1991), pp. 3-15.

[10] J.S. Marron and S.J. Sheather, *Progress in data-based bandwidth selection for kernel density estimation*, Comput. Stat. 11 (1996), pp. 337–381

[11] S. Mallat, *A wavelet tour of signal processing.* Academic Press. New York, 1998 https://doi.org/10.1016/B978-012466606-1/50008-8

[12] F. Palacios-González and R.M. García-Fernández, *A flexible family of density*

Functions, Statistics: A Journal of Theoretical and Applied Statistics, 49 (2014a), pp. 680–704.

https://doi.org/10.1080/02331888.2014.883398

[13] F. Palacios-González and R.M. García-Fernández, *Mixtures of Mixtures Based on Multiresolution Analysis Theory*, Communications in Statistics, Simulation and Computation, 43 (2014b), pp. 723–742 https://doi.org/10.1080/03610918.2012.714031

[14] F. Palacios-González, and R.M. García-Fernández, *A faster algorithm to estimate multiresolution densities*, Computational Statistics, 35 (2020), pp. 1207-1230 https://doi.org/10.1007/s00180-020-00952-w

[15] B.U. Park and J.S. Marron, *Comparison of data-driven bandwidth selectors*. J. Am. Stat. Assoc. 85 (1990), pp. 66–72 https://doi.org/10.1080/01621459.1990.10475307

[16] G. M. Pittau, and R. Zelli, *Testing for changing shapes of income distribution: Italian evidence in the 1990s from kernel estimates*, Empirical Economics, 29 (2004), pp. 415-430. https://doi.org/10.1007/s00181-003-0175-3

[17] G. Schwarz, *Estimating the dimension of a model*, Annals of Statistics, 6 (1978), pp. 461-464.

[18] B.W.Silverman, *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability, vol. 26. Chapman and Hall, London, 1986.

[19] D. W. Scott, R. Tapia and J.R. Thompson, *Kernel density estimation revisited*. Nonlinear Analysis-theory Methods & Applications 1 (1997), pp. 339-372. https://doi.org/10.1016/S0362-546X(97)90003-1

[20] W. Shaoping , L. Ang, W. Kuangyu and W. Ximing, *Robust kernels for kernel density estimation*, Economics Letters, 191 (2020), pp. 109138. doi.org/10.1016/j.econlet.2020.109138.

[21] S.J. Sheather and M.C. Jones, *A reliable data-based bandwidth selection method for kernel density estimation*, J. R. Stat. Soc. Ser. B, 53 (1991), pp. 683–690. https://doi.org/10.1111/j.2517-6161.1991.tb01857.x

[22] S.H. Huang, *Density Estimation by Wavelet-Based Reproducing Kernels*, Statistica Sinica 9 (1999), pp. 137-151.

[23] P. Wojtaszczyk, *A mathematical introduction to wavelets*. Cambridge University Press. London, 1979.

[24] Y. Ziane, S. Adjabi and N. Zougab, *Adaptive Bayesian bandwidth selection in asymmetric kernel density estimation for nonnegative heavy-tailed data*, Journal of Applied Statistics, 42:8 (2015), pp. 1645-1658. American Statistical Association, vol. 91, no. 433, pp. 401–407.

APPENDICES

A.1 CENTRAL AND NON-CENTRAL MOMENTS OF THE CUBIC BOX SPLINE DENSITY

The density function θ has an expected value equal to zero. Hence the central and non-central moments are equal. It is also symmetric and consequently, the odd order moments are null.

PROPOSITION 3

Given $p_1(x)$ and $p_2(x)$ defined by (1), it is proved by polynomial integration that:

$$Q_1(x,r) = \int x^r p_1(2+x)dx = \frac{1}{6}\left[\frac{x^{r+4}}{r+4} + \frac{6x^{r+3}}{r+3} + \frac{12x^{r+2}}{r+2} + \frac{8x^{r+1}}{r+1}\right].$$

Analogously:

$$Q(x,r) = \int x^r p_2(x)dx = \frac{-x^{r+4}}{2(r+4)} + \frac{2x^{r+3}}{r+3} - \frac{2x^{r+2}}{r+2} + \frac{2x^{r+1}}{3(r+1)},$$

and

$$Q_2(x,r) = \int x^r p_2(2+x)dx = \sum_{i=0}^{r}\binom{r}{i}2^{r-i}Q(2+x.i).$$

PROPOSITION 4

The non-central and central moments of order r for θ are[6] zero if r is odd. If r is even then:

$$\mu_\theta(r) \equiv m_\theta(r) = 2\{Q_1(-1,r) - Q_1(2,r) + Q_2(0,r) - Q_2(1,r)\}.$$

Proof. Let us assume that r is even. In this case:

---

[6] The expected value of the density is zero and the central and non-central moments are equal.

$$\mu_\theta(r) = E_\theta[x^r] = \int_{-2}^{2} x^r \theta(x)dx = 2\int_{-2}^{0} x^r \theta(x)dx$$

$$= 2\left\{\int_{-2}^{-1} x^r p_1(2+x)dx + \int_{-1}^{0} x^r p_2(2+x)dx\right\}.$$

Given that θ is a symmetric density with expected value zero, all the central and non-central moments are equal and also the moments with r odd are zero.

Taking into account proposition 1, we can assert that, if r is even, then:

$$\mu_\theta(r) = 2\{Q_1(-1,r) - Q_1(2,r) + Q_2(0,r) - Q_2(1,r)\}.$$

A.2 CENTRAL MOMENTS OF THE DENSITIES $\lambda_{j,k}(x)$

PROPOSITION 5

It is verified that:

$$E_{\lambda_{j,k}}[x] = \frac{k}{s} . \qquad (12)$$

Proof. Let us consider:

$$E_{\lambda_{j,k}}[x] = \int_{-\infty}^{+\infty} x\, \lambda_{j,k}(x)dx = \int_{-\infty}^{+\infty} xs\, \theta(sx-k)dx. \qquad (13)$$

Making the change of variable $y = sx - k$ in (13) we have:

$$E_{\lambda_{j,k}}[x] = \int_{-\infty}^{+\infty} \frac{y+k}{s} s\theta(y)\frac{dy}{s} = \frac{1}{s}\int_{-\infty}^{+\infty} y\,\theta(y)dy + \frac{k}{s}\int_{-\infty}^{+\infty} \theta(y)dy,$$

but $\int_{-\infty}^{+\infty} y\,\theta(y)dy = 0$ y $\int_{-\infty}^{+\infty} \theta(y)dy = 1$.

Hence the proposition is true.

PROPOSITION 6

Let

$$\mu_{j,k}(r) = \int_{-\infty}^{+\infty}\left(x - \frac{k}{s}\right)^r \lambda_{j,k}(x)dx,$$

be the central moment of order r of $\lambda_{j,k}(x)$. It is verified that:

$$\mu_{j,k}(r) = \frac{\mu_{0,0}(r)}{s^r},$$

where $\mu_{0,0}(r) = \mu_\theta(r)$ is, by definition, the moment of order r of the density $\lambda_{0,0}(x) \equiv \theta(x)$ (see proposition 4).

Proof.

$$\mu_{j,k}(r) = \int_{-\infty}^{+\infty} \left(x - \frac{k}{s}\right)^r \lambda_{j,k}(x)dx = \int_{-\infty}^{+\infty} \left(\frac{sx - k}{s}\right)^r \lambda_{j,k}(x)dx$$

$$= \frac{1}{s^r} \int_{-\infty}^{+\infty} (sx - k)^r s\theta(sx - k)dx. \quad (14)$$

Making the change $y = sx - k$ in the integral (14) we have:

$$\mu_{j,k}(r) = \frac{1}{s^r} \int_{-\infty}^{+\infty} y^r \theta(y)dy = \frac{\mu_{0,0}(r)}{s^r}.$$

A.3 CENTRAL MOMENTS OF A MR DENSITY

PROPOSITION 7

Let us consider a MR density as that given by (3). It is verified that:

$$E_f[X] = \int_{-\infty}^{+\infty} xf(x)dx = \sum_{k\in\mathbb{Z}} a_k \frac{k}{s} = \frac{\bar{k}}{s},$$

where:

$$\bar{k} = \sum_{k\in\mathbb{Z}} a_k \times k.$$

Proof. It is trivial taking into account (3).

PROPOSITION 8

It is verified that the central moment of order r of the MR density given by (3) is:

$$\mu_f(r) = \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} \mu_{0,0}(i)\mu_a(r - i),$$

where $\mu_a(r - i) = \sum_{k\in\mathbb{Z}} a_k(k - \bar{k})^{r-i}$ and where $\mu_{0,0}(i) = \mu_\theta(i)$ is calculated according to proposition 4.

Proof. Let:

$$\mu_f(r) = \int_{-\infty}^{+\infty} \left(x - \frac{\bar{k}}{s}\right)^r f(x)dx = \sum_{k\in\mathbb{Z}} a_k \int_{-\infty}^{+\infty} \left(x - \frac{\bar{k}}{s}\right)^r \lambda_{j,k}(x)dx. \quad (15)$$

However,

$$\int_{-\infty}^{+\infty} \left(x - \frac{\bar{k}}{s}\right)^r \lambda_{j,k}(x)dx = \int_{-\infty}^{+\infty} \left(x - \frac{k}{s} + \frac{k}{s} - \frac{\bar{k}}{s}\right)^r \lambda_{j,k}(x)dx. \quad (16)$$

Considering that:

$$\left(x - \frac{k}{s} + \frac{k}{s} - \frac{\bar{k}}{x}\right)^r = \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} (k - \bar{k})^{r-i}(sx - k)^r,$$

and substituting in (14) we have:

$$\int_{-\infty}^{+\infty} \left(x - \frac{\bar{k}}{y} \right)^r \lambda_{j,k}(x)dx = \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} (k - \bar{k})^{r-i} \int_{-\infty}^{+\infty} (sx - x)^i s\theta(sx - k)dx.$$

But making $y = sx - k$ we obtain:

$$\int_{-\infty}^{+\infty} (sx - x)^i s\theta(sx - k) = \int_{-\infty}^{+\infty} (y)^i \theta(y)dy = \mu_\theta(i) \equiv \mu_{0.0}(i).$$

It is verified:

$$\int_{-\infty}^{+\infty} (x - \bar{k})^r \lambda_{j,k}(x)dx = \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} (k - \bar{k})^{r-i} \mu_{0,0}(i).$$

Substituting in (15) we have:

$$\mu_f(r) = \sum_{k \in \mathbb{Z}} a_k \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} (k - \bar{k})^{r-i} \mu_{0,0}(i) = \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} \mu_{0,0}(i) \sum_{k \in \mathbb{Z}} a_k (k - \bar{k})^{r-i}$$

$$= \frac{1}{s^r} \sum_{i=0}^{r} \binom{r}{i} \mu_{0,0}(i) \mu_a(r - i),$$

as we want to prove.

A.4 NON-CENTRAL MOMENTS OF A MR DENSITY

PROPOSITION 9

Taking into account (3) is trivial to prove that:

$$m_j(r) = \sum_{k \in Z} a_k \, m_{\lambda_{j,k}}(r),$$

where

$$m_{\lambda_{j,k}}(r) = \int_{-\infty}^{+\infty} x^r \lambda_{j,k}(x)dx.$$

PROPOSITION 10

It is verified:

$$m_{\lambda_{j,k}}(r) = \sum_{h=0}^{r} \binom{r}{h} m_\theta(r) k^{h-r}$$

where $m_\theta(r) = \mu_\theta(r)$. They are defined and calculated in Proposition 4.

Proof.

$$m_{\lambda_{j,k}}(r) = \int_{-\infty}^{+\infty} x^r \lambda_{j,k}(x)dx = \int_{-\infty}^{+\infty} x^r s\theta(sx - k)dx . \qquad (17)$$

Making the change of variable $y = sx - k$ in (17) it is obtained:

$$\frac{1}{s^r} \int\limits_{-\infty}^{+\infty} (y+k)^r \theta(y) dy. \qquad (18)$$

Taking into account that:

$$(y+k)^r = \sum_{h=0}^{r} \binom{r}{h} y^r k^{h-r}, \qquad (19)$$

and substituting (19) in (18), the proof of the proposition is evident (see proposition 4).

A5. CENTRAL MOMENTS OF A KERNEL DENSITY

PROPOSITION 11

It is verified :

$$E_{\hat{f}_h}[X] = \int\limits_{-\infty}^{+\infty} x\hat{f}_h(x) dx = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} .$$

Proof. Let

$$\int\limits_{-\infty}^{+\infty} x\hat{f}_h(x) dx = \int\limits_{-\infty}^{+\infty} x\frac{1}{nh}\sum_{i=1}^{n} \theta\left(\frac{x-x_i}{h}\right) dx = \frac{1}{nh}\sum_{i=1}^{n}\int\limits_{-\infty}^{+\infty} x\,\theta\left(\frac{x-x_i}{h}\right) dx . \qquad (20)$$

Making the change of variable $y = \frac{x-x_i}{h}$ in the integral (20) we have:

$$\int\limits_{-\infty}^{+\infty} x\,\theta\left(\frac{x-x_i}{h}\right) dx = h\int\limits_{-\infty}^{+\infty} (hy+x_i)\,\theta(y) dy$$

$$= h^2\int\limits_{-\infty}^{+\infty} y\theta(y) dy + hx_i\int\limits_{-\infty}^{+\infty} \theta(y) dy. \qquad (21)$$

But $\int_{-\infty}^{+\infty} y\theta(y) dy = 0$ and $\int_{-\infty}^{+\infty}\theta(y) dy = 1$ so (21) equals hx_i and substituting in (20) we have:

$$E_{\hat{f}_h}[X] = \frac{1}{nh}\sum_{i=1}^{n} hx_i = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} .$$

PROPOSITION 12

It is verified:

$$\mu_{\hat{f}_h}(r) = \sum_{t=0}^{r} \binom{r}{r} h^t m_\theta(t)\, m(r-t).$$

where $m_\theta(t)$ is the non-central moment of order t of density θ . It is obtained following section 2 in the appendix. The expression $\mu(r-t)$ is the sample central moment of order $r-t$. That is:

$$\mu(r-t) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^{t-r}.$$

Proof.

$$\mu_{\hat{f}_h}(r) = \int_{-\infty}^{+\infty}(x - \bar{x})^r \hat{f}_h(x)dx \qquad (22)$$

That is:

$$\mu_{\hat{f}_h}(r) = \frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}(x - \bar{x})^r \theta\left(\frac{x - x_i}{h}\right)dx. \quad (23)$$

Making the change of variable $y = \frac{x-x_i}{h}$ in (23) we have:

$$\int_{-\infty}^{+\infty}(x - \bar{x})^r \theta\left(\frac{x - x_i}{h}\right)dx = h\int_{-\infty}^{+\infty}(hy + x_i - \bar{x})^r \theta(y)dy. \quad (24)$$

But

$$(hy + x_i - \bar{x})^r = \sum_{t=0}^{r}\binom{r}{t}h^t y^t (x_i - \bar{x})^{r-t}.$$

Substituting in (24) we have:

$$h\int_{-\infty}^{+\infty}(x - \bar{x})^r \theta\left(\frac{x - x_i}{h}\right)dx = \sum_{t=0}^{r}\binom{r}{t}h^{t+1}m_\theta(t)(x_i - \bar{x})^{r-t}.$$

Substituting the latter expression in (23) we obtain:

$$\mu_{\hat{f}_h}(r) = \frac{1}{nh}\sum_{i=1}^{n}\sum_{t=0}^{r}\binom{r}{t}h^{t+1}m_\theta(t)(x_i - \bar{x})^{r-t} = \sum_{t=0}^{r}\binom{r}{t}h^t m_\theta(t)\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^{r-t}.$$

Naming

$$\mu(r-t) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^{r-t},$$

the proposition is proven.

A.6. ASYMPTOTIC PROPERTIES OF THE MR MOMENTS

Proof of Proposition 1.

By definition (see [12,13]):

$$\hat{a}_k = \frac{1}{n}Card\left\{x_i\ i = 1, 2, \dots, n\ \Big|\ x_i \in \left[\frac{k-2}{2^j}, \frac{k+2}{2^j}\right)\right\}.$$

It is evident that $x_i \in \left[\frac{k-2}{2^j}, \frac{k+2}{2^j}\right)$ if and only if $k = k(x_i)$.

Moreover, the intervals:

$$\left[\frac{k-2}{2^j}, \frac{k+2}{2^j}\right),$$

have center $\frac{k}{2^j}$ and radius $\frac{1}{2^{j-1}}$, so for j large enough the previous intervals will have a radius so small that each sample element, x_i, belongs to a different interval. In this case, the coefficients greater than zero are those associated with intervals that contain a sample element, that is:

$$\hat{a}_{k(x_i)} = \frac{1}{n} \quad i = 1, 2, \dots, n.$$

Hence the proposition is true.

Proof of Proposition 2.

$$\hat{m}_r = \int_{-\infty}^{+\infty} x^r \hat{f}_j(x)\,dx = \int_{-\infty}^{+\infty} x^r \sum_{k \in Z} \hat{a}_k \lambda_{j,k}(x)\,dx = \sum_{k \in Z} \hat{a}_k \int_{-\infty}^{+\infty} x^r \lambda_{j,k}(x)\,dx.$$

For a j large enough, according to proposition 1, we have:

$$\hat{a}_k = \begin{cases} 0 & if\ k \neq k(x_i)\ \forall i = 1, 2, \dots, n \\ \dfrac{1}{n} & if\ k = k(x_i)\ \forall i = 1, 2, \dots, n \end{cases},$$

which allows us to write:

$$\hat{m}_r = \frac{1}{n} \sum_{i=1}^{n} \int_{-\infty}^{+\infty} x^r \lambda_{j,k(x_i)}(x)\,dx. \qquad (25)$$

However,

$$\int_{-\infty}^{+\infty} x^r \lambda_{j,k(x_i)}(x)\,dx = \int_{-2}^{+2} x^r s\theta\big(sx - k(x_i)\big)\,dx, \qquad (26)$$

with $s = 2^j$.

If we make the change of variable $y = sx - k(x_i)$ in (25), we have:

$$\int_{-2}^{+2} x^r s\theta\big(sx - k(x_i)\big)\,dx = \int_{-2}^{+2} \left(\frac{y + k(x_i)}{s}\right)^r \theta(y)\,dy. \qquad (27)$$

If j tends to infinite, we have:

$$\lim_{j\to\infty} \frac{y}{s} = 0 \qquad\qquad \lim_{j\to\infty} \frac{k(x_i)}{s} = x_i. \qquad (28)$$

The first equality in (28) is trivially true and the second is also true since by definition:

$$x_i \in \left[\frac{k(x_i) - 2}{s}, \frac{k(x_i) - 2}{s}\right),$$

and when j tends to infinity, the radius of the interval converges to zero and its center is the point $\frac{k(x_i)}{s}$.

Assuming that (28) is true we can write:

$$\lim_{j\to\infty} \int_{-2}^{+2} \left(\frac{y + k(x_i)}{s}\right)^r \theta(y)dy = \int_{-2}^{+2} (x_i)^r \theta(y)dy = x_i^r \int_{-2}^{+2} \theta(y)dy = x_i^r . \quad (29)$$

Wherewith, under (25) and (29), we have:

$$\lim_{j\to\infty} \widehat{m}_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r .$$

That is, the non-central moments of the estimated MRDE converge to the non-central moments of the sample.