Density Estimation & Nonparametric Regression

Urban Economics

Ignacio Sarmiento-Barbieri

Universidad de los Andes

February 15, 2022

Agenda

1. Recap: Univariate Density Estimation
2. Kernel Density
3. Regressions: Non parametric estimation of conditional expectations
   - Local Constant Kernel Estimation
   - Local Linear Kernel Estimation
4. Further Readings

1 Recap: Univariate Density Estimation

2 Kernel Density

3 Regressions: Non parametric estimation of conditional expectations
- Local Constant Kernel Estimation
- Local Linear Kernel Estimation

4 Further Readings

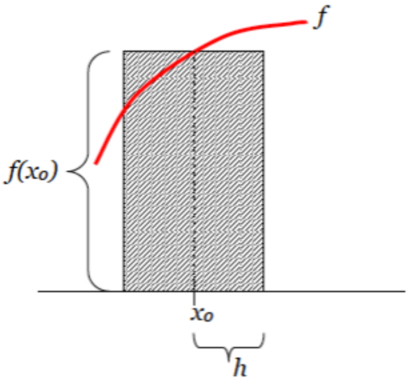
Univariate density estimation: parametric vs nonparametric methods

- Parametric Methods
- Non paraetric
  - Histogram
  - Naive

Another non-parametric estimator: the naive estimator

Continuous case

- X is a continuous RV ($Pr\left(X = x0\right) = 0$) we evaluate the probability that X is "close" to x_0.
- We say that x_i is close to x_0 if x_i belongs to the interval $\left(x_o - \ h, \ x_o + \ h\right)$



The naive estimator

- Then:

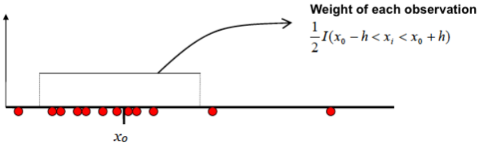
$$f(x_o) \cong \frac{1}{2h \times n} \sum_{i=1}^{n} I\left(-1 < \frac{x_i - x_o}{h} < 1\right) \tag{1}$$

- $\hat{f}(x_o)$ is an estimator of the height of the rectangle in the previous graph

The naive estimator

Notice:

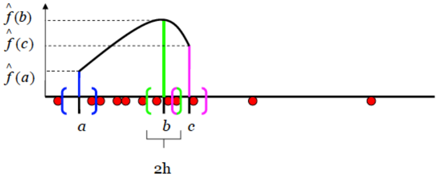
- Every possible point x_o is the center of an interval
- Observations that are within that interval (less than one h away from x_o) are weighted by 1/2



The naive estimator

Notice:

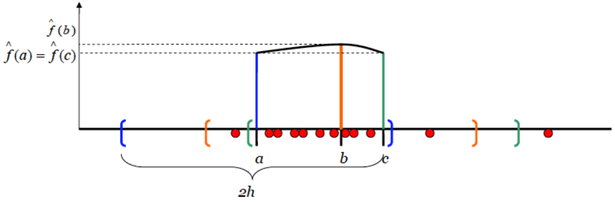
- To estimate the density at points a, b and c we construct intervals around each
- Then, unlike what happens in the histograms, the intervals of the naive estimator overlap



The naive estimator

Problems

- It depends on h: the larger the bandwidth, the more distant observations from x_o are used to estimate f(x_o).The higher h, the smoother the estimated density (h=smoothing parameter)



- The weight $1/2I(.)$ is discontinuous at the limits of each interval, generating discontinuities in the estimated density.
- The weight treats observations very close to x_o in the same way as others somewhat further away, as long as they belong to the interval of length 2h around x_o

1 Recap: Univariate Density Estimation

2 Kernel Density

3 Regressions: Non parametric estimation of conditional expectations
- Local Constant Kernel Estimation
- Local Linear Kernel Estimation

4 Further Readings

The weighted average estimator or kernel method

- The kernel estimator is a generalization of the naive estimator that overcomes some of the deficiencies of the latter.
- The weight $1/2I(.)$ of the naive estimator is replaced by a new weight K(.):

$$\hat{f}\,(x_o) = \frac{1}{n\,\times h}\sum_{i=1}^{n} K\left(\frac{x_i - x_0}{h}\right) \tag{2}$$

- Rosenblatt-Parzen density estimation approach

The weighted average estimator or kernel method

- The weights or kernels are known functions that satisfy:

  1. $K(\phi) \geq 0$
  2. $\int_{-\infty}^{\infty} K\ (\phi)\ d\phi = 1$
  3. $K(\phi) = K(-\phi)$

The weighted average estimator or kernel method

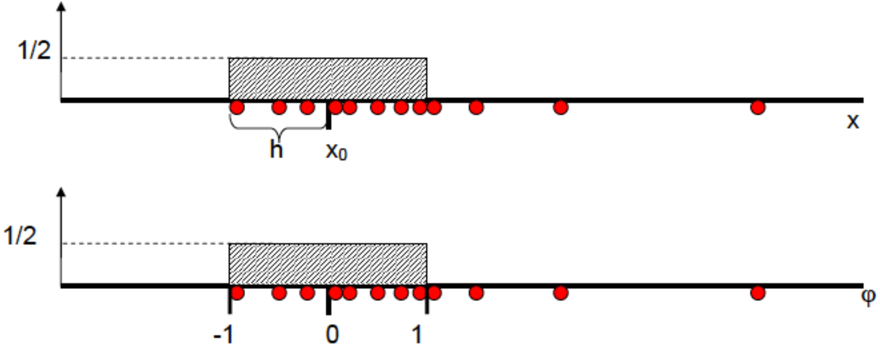
Example 1: rectangular or uniform kernel

$$K(\phi) = \begin{cases} \frac{1}{2} & |\phi| < 1 \\ 0 & o.w. \end{cases} \tag{3}$$

where $\phi = \frac{x_i - x_0}{h}$ In this case, the Kernels estimator matches the naive estimator. Note that the weight is similar to a uniform density function on the interval (-1, 1)

The weighted average estimator or kernel method

Example 1: rectangular or uniform kernel



The weighted average estimator or kernel method

Example 2: Gaussian kernel

- Gaussian

$$K(\phi) = \frac{1}{\sqrt{2\pi}} e^{-\phi^2/2} \tag{4}$$

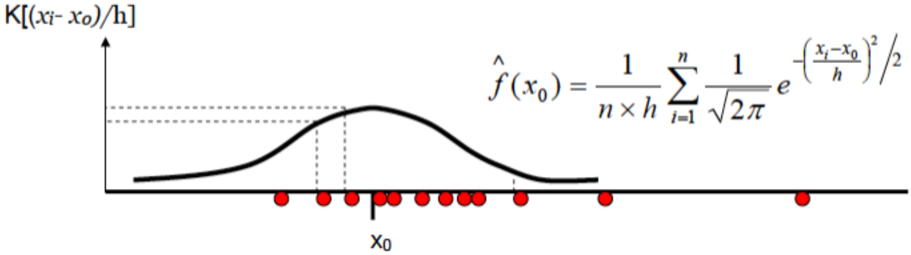
- The kernel function corresponds to the standard normal density function that satisfies the above assumptions.
- For this type of kernel, the density estimator is given by:

$$\hat{f}\ (x_o) = \frac{1}{n\ \times h} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\ \left(\frac{x_i - x_o}{h}\right)^2/2} \tag{5}$$

- Important: we don't assume normal distribution for any variable. The functional form is just used to weight the sample observations

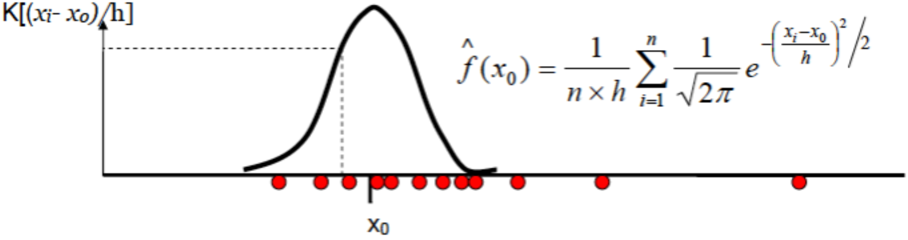
The weighted average estimator or kernel method

Example 2: Gaussian kernel



The weighted average estimator or kernel method

Example 2: Gaussian kernel (smaller h)



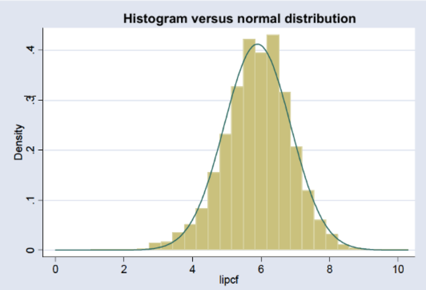
The weighted average estimator or kernel method

Other Kernels

Kernel type	Formula	Support
Gaussian or normal	$\frac{1}{\sqrt{2\pi}}\exp(-\varphi^2/2)$	$\varphi \in (-\infty, \infty)$
Epanechnikov	$\frac{3}{4\sqrt{5}}(1-\frac{1}{5}\varphi^2)I(\lvert\varphi\rvert<\sqrt{5})$	$\varphi \in \left(-\sqrt{5}, \sqrt{5}\right)$
Epanechnikov modificado	$\frac{3}{4}(1-\varphi^2)I(\lvert\varphi\rvert<1)$	$\varphi \in (-1, 1)$
Triangular	$(1-\lvert\varphi\rvert)I(\lvert\varphi\rvert<1)$	$\varphi \in (-1, 1)$
Uniform or rectangular	$\frac{1}{2}I(\lvert\varphi\rvert<1)$	$\varphi \in (-1, 1)$

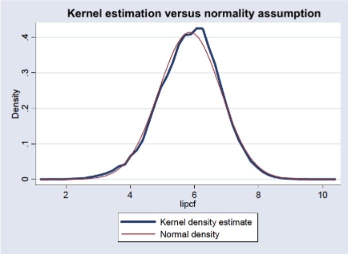
The weighted average estimator or kernel method

Example: Income Distribution



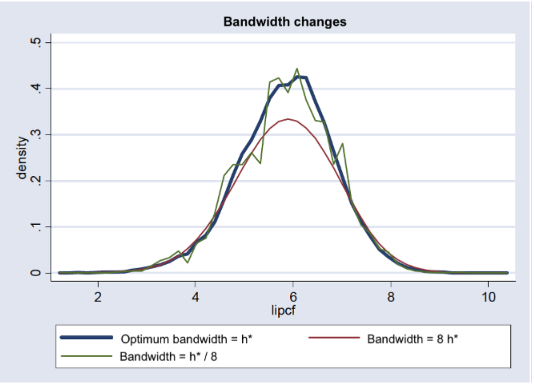
The weighted average estimator or kernel method

Example: Income Distribution



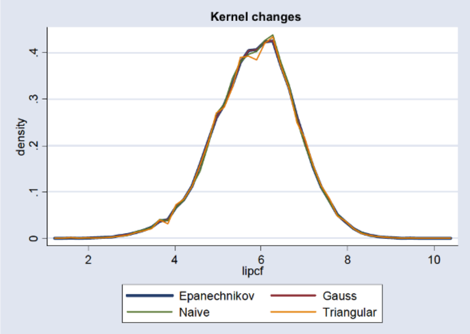
The weighted average estimator or kernel method

Example: Income Distribution



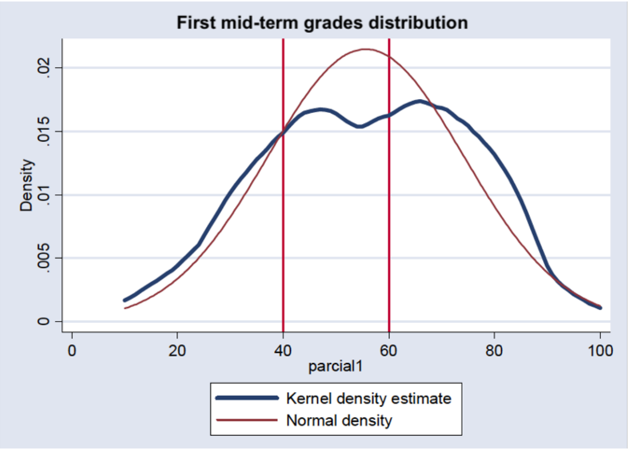
The weighted average estimator or kernel method

Example: Income Distribution



The weighted average estimator or kernel method

Example: Income Distribution



The weighted average estimator or kernel method

Properties: Bias

- The kernel estimator is generally biased.
- The approximate expression for the bias is given by:

$$bias[\hat{f}(x_o)] \approx \frac{h^2}{2} f''(x_o) \int_{-\infty}^{\infty} K(\phi)\phi^2 d\phi \tag{6}$$

The weighted average estimator or kernel method

Properties: Variance

- The approximate expression for the asymptotic variance of the kernel estimator is given by:

$$variance[\hat{f}(x_o)] \approx \frac{1}{n \times h} f(x_o) \int_{-\infty}^{\infty} K^2(\phi) d\phi \tag{7}$$

The weighted average estimator or kernel method

Properties: MSE

$$MSE(\hat{f}(x)) = Var(\hat{f}(x)) + \left(bias(\hat{f}(x))\right)^2 \tag{8}$$

The weighted average estimator or kernel method

Properties: MSE

$$MSE(\hat{f}(x)) = Var(\hat{f}(x)) + \left(bias(\hat{f}(x))\right)^2 \tag{8}$$

$$IMSE(\hat{f}(x)) = \int MSE(\hat{f}(x))dx \tag{9}$$

The weighted average estimator or kernel method

Bandwidth Selection

- We obtain a bandwidth which globally balances bias and variance by minimizing IMSE with respect to h, i.e.,

$$h_{opt} = \left(\frac{\int K^2(z)dz}{(\int z^2 K(z)dz)^2 \int f'(x)^2 dx} \right)^{-1/5} n^{-1/5} \tag{10}$$

The weighted average estimator or kernel method

Bandwidth Selection

- There are four general approaches to bandwidth selection,
  1. reference rules-of-thumb,

The weighted average estimator or kernel method

Reference Rule-of-Thumb

- The reference rule-of-thumb for choosing the bandwidth uses a standard family of distributions
- Under a gaussian density and kernel

$$h_{opt} = 1.059\sigma n^{-1/5} \tag{11}$$

The weighted average estimator or kernel method

Bandwidth Selection

- There are four general approaches to bandwidth selection,

  1. reference rules-of-thumb,
  2. plug-in methods,
  3. cross-validation methods,
  4. bootstrap methods.

Multivariate Density

- Suppose we have X and Y iid
- We want to estimate

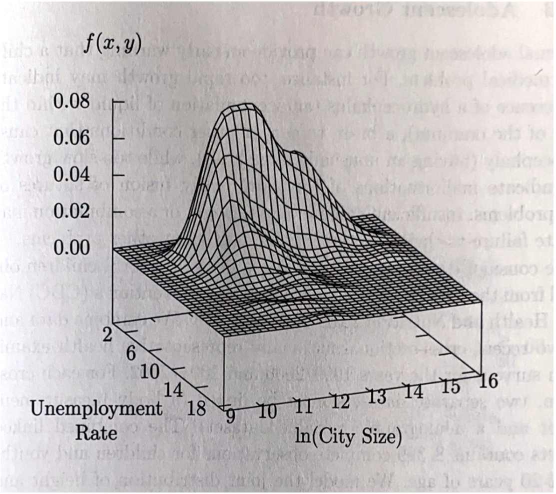
$$f(x, y) \tag{12}$$

Multivariate Density

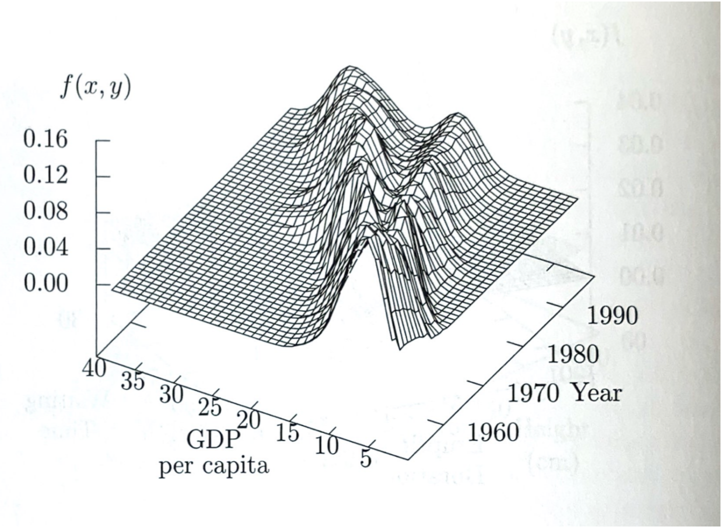
- We use a product kernel function

$$\hat{f}(x,y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{Y_i - x}{h_x}\right) \tag{13}$$

Multivariate Density



Multivariate Density



1 Recap: Univariate Density Estimation

2 Kernel Density

3 Regressions: Non parametric estimation of conditional expectations
- Local Constant Kernel Estimation
- Local Linear Kernel Estimation

4 Further Readings

Non parametric estimation of conditional expectations

- In this part we are interested in the relationship between y and x where y is the dependent variable and x is the explanatory variable.

$$y = f(x) + u \tag{14}$$

- where we make no assumptions about $m(x)$
- The only assumption we add is $E(u|x) = 0$.

Non parametric estimation of conditional expectations

$$E(y|x) = m(x) \tag{15}$$

- This is true if $E(u|x) = 0$.
- The question is how we estimate this $m(x)$

Local Constant Kernel Estimation (Nadaraya–Watson)

$$E(y|x) = \int_{-\infty}^{\infty} yf(y|x)dy$$

Local Constant Kernel Estimation (Nadaraya–Watson)

$$
\begin{aligned}
E(y|x) &= \int_{-\infty}^{\infty} yf(y|x)dy \\
&= \int_{-\infty}^{\infty} y\frac{f(y,x)}{f(x)}dy \\
&= \frac{\int_{-\infty}^{\infty} yf(y,x)dy}{f(x)} \\
&= \frac{v(x)}{g(x)}
\end{aligned}
$$

Local Constant Kernel Estimation (Nadaraya–Watson)

Then the estimation we can propose is

$$\hat{E}(y|x) = \hat{m}(x) = \frac{\int_{-\infty}^{\infty} y\hat{f}_{x,y}(y,x)dy}{\hat{f}_x(x)} = \frac{\hat{v}(x)}{\hat{g}(x)}$$

then

$$\hat{f}_x(x) = \frac{1}{nh^q}\sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)$$

where $\underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right) = K\left(\frac{x_{1i}-x}{h}\right).K\left(\frac{x_{2i}-x}{h}\right)......K\left(\frac{x_{qi}-x}{h}\right)$

Local Constant Kernel Estimation (Nadaraya–Watson)

$$\hat{f}_{x,y}(y,x) = \frac{1}{nh^{q+1}} \sum K\left(\frac{y_i - y}{h}\right) \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)$$

replacing

$$\hat{v}(x) = \frac{1}{nh^{q+1}} \sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right) \int yK\left(\frac{y_i - y}{h}\right) dy$$

$$\hat{v}(x) = \frac{1}{nh^{q}} \sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right) y_i$$

Local Constant Kernel Estimation (Nadaraya–Watson)

Then we have

$$\hat{m}(x) = \frac{\hat{v}(x)}{\hat{g}(x)} = \frac{\frac{1}{nh^q}\sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right) y_i}{\frac{1}{nh^q}\sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)} = \frac{\sum y_i \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)}{\sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)}$$

Note that we can write $w_i = \frac{\underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)}{\sum \underline{\mathrm{K}}\left(\frac{x_i - x}{h}\right)}$

$$\hat{m}(x) = \sum y_i w_i$$

Local Constant Kernel Estimation (Nadaraya–Watson)

We can show that this is the same to solving the followign problem

$$\underset{a}{Min} \sum_{i=1}^{n} w_i (y_i - a)^2$$

where $\hat{a}(x) = \hat{m}(x)$.

1 Recap: Univariate Density Estimation

2 Kernel Density

3 Regressions: Non parametric estimation of conditional expectations
- Local Constant Kernel Estimation
- Local Linear Kernel Estimation

4 Further Readings

Local Linear Kernel Estimation

We can think to fit a line instead of a constant

$$\underset{\{a,b\}}{Min} \sum_{i=1}^{n} w_i \left\{ y_i - \left[a + b \left(x_i - x \right) \right] \right\}^2$$

Intuition

- The question is how we can write $m(x_i)$ in terms of x.

Intuition

- The question is how we can write $m(x_i)$ in terms of x.
- With a linear approximation

$$m(x_i) \approxeq \underbrace{m(x)}_{a} + \underbrace{m'(x)}_{b}(x_i - x)$$

- We get a marginal effect

Local Linear Kernel Estimation

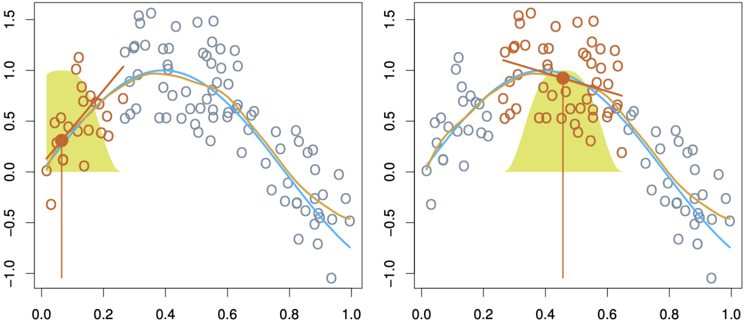
We can show that

$$\hat{a}(x) = y_w - \hat{b} \quad \underbrace{(\bar{x}_w - x)}_{\text{weighted local mean}}$$

$$\hat{b}(x) = \frac{\sum (y_i - \bar{y}_w)(x_i - \bar{x}_w) w_i}{\sum (x_i - \bar{x}_w) w_i}$$

with $\bar{x}_w = \frac{\sum x_i w_i}{\sum w_i}$ y $\bar{y}_w = \frac{\sum y_i w_i}{\sum w_i}$ are the weighted means

Intuition



Example: Geographical Weighted Regressions

The determinant of educational achievement in Georgia (Fotheringham et al., 2002)

$$Bachelor\,or\,higher\,degrees_{county} = f(Rural, Eld, FB, Pov, Black) + u \tag{16}$$

Example: Geographical Weighted Regressions

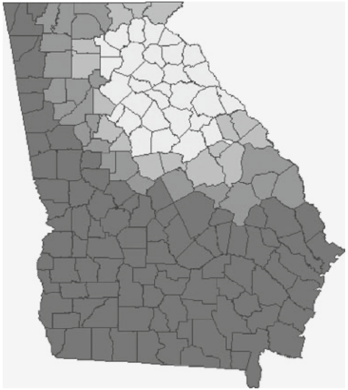
The determinant of educational achievement in Georgia (Fotheringham et al., 2002)

$$Bachelor\,or\,higher\,degrees_{county} = f(Rural, Eld, FB, Pov, Black) + u \tag{16}$$

Parameter	*Minimum*	*First Quartile*	*Median*	*Third Quartile*	*Maximum*	*Range*	*Global (OLS)*
Intercept	14.170000	15.350000	17.050000	18.200000	18.860000	4.69	17.2437
Rural	−0.081350	−0.073480	−0.064850	−0.055110	−0.051080	0.03027	−0.0703
Eld	−0.191200	−0.094630	−0.065330	−0.032360	0.012500	0.2037	0.0114
FB	0.854300	1.282000	2.031000	2.796000	3.138000	2.2837	1.8525
Pov	−0.304800	−0.258100	−0.196100	−0.115100	−0.034210	0.27059	−0.2552
Black	−0.016900	0.006347	0.031610	0.060620	0.087210	0.07031	0.0491

Example: Geographical Weighted Regressions

The determinant of educational achievement in Georgia (Fotheringham et al., 2002)



Further Readings

- Li, Q., & Racine, J. S. (2007). Nonparametric econometrics: theory and practice. Princeton University Press.
- Pagan, A. & Ullah (1999). Nonparametric Econometrics. Cambridge University Press.
- Silverman (1998). Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC.