ECON E572
Department of Economics

Yoosoon Chang
Spring 2020

Part IV: Problem Set 1

Due on April 27 Monday

Use the cross-sectional data on happiness (y_i), age (x_i), and also income (w_i) as one of the covariates relevant for explaining happiness to answer the following questions. The data along with a detailed description on the data source and on how they are constructed will be posted on the Canvas. Obtain your answers using your own code written in Matlab. You are encouraged to work with your classmates, but I expect that you write your solutions independently.

1. Answer the following:

(a) For each (x_i) and (y_i), obtain the histogram.

(b) For each (x_i) and (y_i), obtain the naive density estimator.

(c) For both (x_i) and (y_i), obtain the Kernel density estimator with four different Kernel functions - Gaussian, Bartlett, Epanechnikov, and Parzen.

2. Let $\mathbb{E}(y_i|x_i) = m(x_i)$, and we consider the nonparametric regression model

$$y_i = m(x_i) + \varepsilon_i$$

where (y_i, x_i) are iid observations, $i = 1, ..., n$, and $\mathbb{E}(\varepsilon_i|x_i) = 0$. Answer the following:

(a) Estimate the regression function m using the local constant estimation method with the four Kernel functions used to answer Part (c) in Question 1 above.

(b) Compare the regression function estimate $\hat{m}$ that you obtained nonparametrically from Part (a) with the estimate that you may obtain from the linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$. Discuss the results implied by the two regression functions estimates.

(c) Obtain the local linear estimator $\tilde{m}(x)$ of the regression function $m(x)$ using the bandwidth h_n set by the rule of thumb, viz., $h_n(x) = \hat{\sigma}_n n^{-1/5}$, where $\hat{\sigma}_n$ is the standard error of x_i's, and with the four kernel functions.

(d) Select the bandwidth parameter h_n^{cv} by minimizing the cross-validation function

$$CV(h) = \sum_{i=1}^{n} \left(\hat{m}_{-1}(x_i) - y_i\right)^2$$

where $\hat{m}_{-1}(x_i)$ is the leave-one-out estimate of $m(x_i)$ given by

$$\hat{m}_{-1}(x_i) = \frac{\sum_{j\neq i}^{n} K\left(\frac{x_j - x_i}{h}\right) y_i}{\sum_{j\neq i}^{n} K\left(\frac{x_j - x_i}{h}\right)}$$

Use this bandwidth h_n^{cv} to obtain the local linear estimator $\tilde{m}^{cv}(x)$ with the same kernel functions considered in Part (c). Compare $\tilde{m}^{cv}(x)$ with $\tilde{m}(x)$ obtained in Part (c).

3. Let $\mathbb{E}(y_i|w_i, x_i) = w_i\beta + m(x_i)$, where w_i signifies a control variable, and consider the partially linear regression model

$$y_i = w_i\beta + m(x_i) + \varepsilon_i$$

where (y_i, w_i, x_i) are iid observations, $i = 1, ..., n$, and $\mathbb{E}(\varepsilon_i|w_i, x_i) = 0$. Obtain the answers to the following questions using the bandwidth parameters chosen by the rule-of-thumb method and the cross-validation approach, and with the four kernel functions introduced above.

(a) Show that we may rewrite the above regression as

$$y_i - \mathbb{E}(y_i|x_i) = (w_i - \mathbb{E}(w_i|x_i))\,\beta + \varepsilon_i$$

(b) Estimate the conditional expectations $\mathbb{E}(y_i|x_i)$ and $\mathbb{E}(w_i|x_i)$ by the local constant estimation method using the observations (y_i, x_i) and (w_i, x_i), respectively.

(c) Estimate the parameter β in the linear part of the given partially linear model by the OLS estimator $\hat{\beta}$ from the following linear regression

$$y_i - \widehat{\mathbb{E}(y_i|x_i)} = \left(w_i - \widehat{\mathbb{E}(w_i|x_i)}\right)\beta + \varepsilon_i$$

where $\widehat{\mathbb{E}(y_i|x_i)}$ and $\widehat{\mathbb{E}(w_i|x_i)}$ denote the local constant estimators obtained in Part (b).

(d) Plug $\hat{\beta}$ into the given model as

$$y_i - w_i\hat{\beta} = m(x_i) + \varepsilon_i$$

and estimate the nonparametric part $m(x)$ of the regression function from here by the local constant estimator. Denote the resulting estimator by $\hat{m}^{PL}(x)$.

(e) Compare $\hat{m}^{PL}(x)$ with the local constant estimator $\hat{m}(x)$ obtained from the purely nonparametric regression $y_i = m(x_i) + \varepsilon_i$ considered in Question 2.