Problem Set 3

Eco 523 - Nonparametric Econometrics

Carolina Caetano

The data sets used in this problem set can be found in the website: http://www.wadsworth.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9780324581621&discipline_number=413&token=. Click on the "Data Sets" link on the left.

1 Theoretical Part

Question 1.1:

(i) Generate a 4^{th} and a 6^{th} order kernel from the uniform and triangular kernels.

(ii) Generate a 4^{th} order kernel from the epanechnikov kernel in such a way that no observation in the sample is weighted with a negative weight by the kernel. (This shows how in finite samples we can use regular second order kernels without loss of generality).

(iii) Show that the bias of the Nadaraya-Watson estimator is negligible when we use a 4^{th} order kernel. Can we say that the bias is zero?

Question 3: Describe in detail how you would perform the cross-validation method in order to choose the bandwidth in the estimation of the non-parametric term in the partially linear estimator.

Question 3:

(i) Suggest a local polynomial estimator of the single-index model:

$$Y = g(X^T\beta) + u.$$

(ii) Suggest an estimator of $\dfrac{\partial E(Y|X=x)}{\partial x}$.

Question 4: Suggest an estimator for the function g below (an additive partially linear model):

$$Y = g(X, Z, Q) = \beta X + g_1(Z) + g_2(Q) + u$$

You should be able to do this by combining the techniques of the partially linear model with the additive model.

Question 5:

(i) In the average treatment effect model in the notes, suppose that there are two treatment levels. That means that we have three groups: the control ($T = 0$), the first group ($T = 1$), where a level of treatment is applied, and the second group ($T = 2$), where a different level of treatment is applied. Can you suggest an estimator of the average effect of treatment?

(ii) How would the regression discontinuity design estimator change if the treatment was applied to all observations such that $X \geqslant \bar{x}$?

Question 6: Suppose that in the regression discontinuity design, the assumption

$$\mathbb{E}(u_1|X = \bar{x}) = \mathbb{E}(u_0|X = \bar{x}).$$

is substituted by

$$\mathbb{E}(u_1|X = \bar{x}, Z) = \mathbb{E}(u_0|X = \bar{x}, Z).$$

for a set of observable covariates Z. This means that we cannot in fact believe that there is random assignment around $\bar{x}$, but rather random assignment conditional on covariates. Suppose that Z can assume M values: $\{z_1, \ldots, z^m\}$. How would you estimate θ, the local average treatment effect at $\bar{x}$?

Question 6: Let the model be

$$Y = g(X) + u.$$

where X is scalar and endogenous, and Z is a continuously distributed variable that is dependent of X and independent of u. Describe how you would test the following hypothesis:

$$H_0 : \mathbb{E}(g'(X)) = 0.$$

2 Empirical Part

Question 2.1: (This question is about partial linear regression)

Use the data in SLEEP75.RAW from Biddle and Hamermesh (1990). Consider the model

$$totwrk = g(sleep) + \beta_1\, educ + \beta_2\, age + u$$

where *educ* and *age* are education and age measured in years respectively.

(i) Estimate β_1 and β_2 and test whether $\beta_1 = 0$. In order to build the plugins, use a spline basis with degree 3 and smoothness 2, and the knot structure you used in question 2.2 (iii) in problem set 2.

(ii) Estimate $g(2940)$ by kernel regression with the bandwidth used in question 5 (ii) in problem set 1.

Question 2.2: (This question is about the varying coefficients model)

Use the same data set from the last question to study the model

$$totwrk = educ\,\beta(sleep) + u.$$

Estimate the expected difference in work per week if a person which studied for 13 years and sleeps 8 hours per day starts sleeping 7 hours per day instead. Use a spline basis with degree 3 and smoothness 2, and the knot structure you used in question 2.1 (i).