Applied Quantitative Method (II)

Department of Economics

National Taipei University

Spring Semester

Homework 3

(Due in Class on the Date Assigned)

Empirical Problem Set: This homework assignment is from *http://www.the-smooth-operators.com/* (with some major revisions). Answer the following questions using the data *cps71* from the np package in R.

Heckman and Polachek (1974) suggest a quadratic parametric relationship between earnings and age

$$y_i = \alpha + \beta z_i + \gamma x_i + \delta x_i^2 + u_i,$$

where y_i is the logarithm of earnings, z_i is education and x_i is age. Mincer (1974) finds that earnings increase with age through much of the working life but the rate of increase diminishes with age. Pagan and Ullah (1999) present a local-constant kernel estimate of an age earnings profile based on Canadian data (cps71 - available in the np package in R) for n=205 males having common education (high school)

$$y_i = m(\overline{z}, x_i) + u_i.$$

1. Using the Li and Wang (1999) test, test that the quadratic parametric specificiation is appropriate (use a wild bootstrap). Use (1) the rule-of-thumb bandwidth and (2) the cross-validated bandwidth from the local-constant estimator with the Gaussian kernel.

2. Are your results consistent with your answers to the question 3 from the previous homework assignment?

3. A random variable v following a standard normal distribution is included in the model as the second regressor. Using the Lavergne and Vuong (2000) test, test that v is an irrelevant regressor (use a wild bootstrap). Use (1) the rule-of-thumb bandwidth and (2) the cross-validated bandwidth from the local-constant estimator with the Gaussian kernel.

4. Is the cross-validated bandwidth greater than two times the standard deviation of v? What does this mean? Is this consistent with the result of the hypothesis test?

Theoretical Problem Set: Answer the following questions about testing in kernel regression.

1. Several kernel-based test statistics require exclusion of centering terms in a double summation ($\sum_{j=1}^{n}\sum_{i=1,i\neq j}^{n}$) because centering terms ($i = j$), which lead to biases, do not disappear as the sample size goes to infinity. Asymptotically centering terms disappear because kernels become zero for $i \neq j$ but not for $i = j$. Based on the key assumption under which consistency can be satisfied, explain why kernels change in that way.

2. Bootstrap is a re-sampling method and it can be used to construct a reasonable finite-sample distribution of a kernel-based test statisitc, such as wild bootstrap. Show that the following hold for the wild bootstrapped residuals: $E[u_i^*] = 0$ and $E[(u_i^*)^2] = \widehat{u}_i^2$.