**Applied Quantitative Method (II)**

Department of Economics

National Taipei University

Spring Semester

**Homework 2**

(Due in Class on the Date Assigned)

**Empirical Problem Set:** This homework assignment is from *http://www.the-smooth-operators.com/* (with some minor 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. Plot local-constant and local-linear estimates along with their error bounds (use a wild bootstrap) using a standard normal kernel with (1) rule of thumb bandwidth and (2) bandwidth calculated using least-squares cross-validation.

2. Is the dip present in the resulting nonparametric estimates? How do Pagan and Ullah (1999) explain the dip (see the chapter 3.14.2)?

3. Without conducting a formal test, does the dip appear to be significant?

4. Which nonparametric estimator appears to provide the most appropriate fit to this data?

5. Based on your answer to the previous question, plot its gradients along with their error bounds (use a wild bootstrap).

**Theoretical Problem Set:** The following is an optimization problem about the local linear estimator.

$$\min_{m(x),\beta(x)} \sum_{i=1}^{n} [y_i - m(x) - (x_i - x)\beta(x)]^2 k\left(\frac{x_i - x}{h}\right)$$

1. Derive the local linear estimator $m(x)$ and its gradient $\beta(x)$.

2. Is the gradient $\beta(x)$ the same as the least-squares estimator of the coefficient $\beta$ from a simple linear regression model $y = \alpha + \beta x + u$ when the bandwidth goes to infinity?