APEC 8213: Econometric Analysis III
Spring 2026
Professor: Paul Glewwe
Teaching Assistant: Matt Braaksma

Problem Set 3

Handed out: February 9, 2026
Due: February 18, 2026

1. Tobit and CLAD Estimation

(30 points total)

Load the CHJ2004 dataset. The variables `tinkind` and `income` are household transfers received in-kind and household income, respectively. Divide both variables by 1000 to standardize. Create the regressor Dincome $= (\text{income} - 1) \times \mathbb{1}\{\text{income} > 1\}$.

(a) Estimate a linear regression of `tinkind` on `income` and `Dincome`. Interpret the results. **(5 points)**

(b) Calculate the percentage of censored observations (the percentage for which `tinkind` = 0). Do you expect censoring bias to be a problem in this example? **(5 points)**

(c) Suppose you try and fix the problem by omitting the censored observations. Estimate the regression on the subsample of observations for which `tinkind` > 0. **(5 points)**

(d) Estimate a Tobit regression of `tinkind` on `income` and `Dincome`. **(5 points)**

(e) Estimate the same regression using CLAD. **(5 points)**

(f) Interpret and explain the differences between your results in (a)-(e). **(5 points)**

2. Censored Outcomes

(30 points total)

Load the DDK2011 dataset. Create a variable `testscore` which is `totalscore` standardized to have mean zero and variance one. The variable `tracking` is a dummy indicating that the students were assigned to different classes based on initial test scores. The variable `percentile` is the student's percentile in the initial distribution. For the following regressions, cluster by school.

(a) Estimate a linear regression of `testscore` on `tracking`, `percentile`, and percentile^2. Interpret the results. **(8 points)**

(b) Suppose the scores were censored from below. Create a variable `ctest` which is `testscore` censored at 0. Estimate a linear regression of `ctest` on `tracking`, `percentile`, and percentile^2. How would you interpret these results if you were unaware that the dependent variable was censored? **(8 points)**

(c) Suppose you try and fix the problem by omitting the censored observations. Estimate the regression on the subsample of observations for which `ctest` is positive. **(7 points)**

(d) Interpret and explain the differences between your results in (a), (b), and (c). **(7 points)**

3. Silverman's Optimal Bandwidth Rule of Thumb

(40 points total)

Recall that the asymptotically optimal bandwidth for kernel density estimation, obtained by minimizing the asymptotic integrated mean squared error (AIMSE), is

$$h_0 = \left(\frac{R(K)}{\mu_2(K)^2 R(f'')} \right)^{1/5} n^{-1/5}, \tag{17.11}$$

where

$$R(K) = \int K(u)^2 \, du, \quad \mu_2(K) = \int u^2 K(u) \, du, \quad R(f'') = \int (f''(x))^2 \, dx.$$

The quantity $R(f'')$ is unknown in practice.

Assume that the data are generated from a normal distribution $X \sim \mathcal{N}(0, \sigma^2)$. Derive **Silverman's rule-of-thumb bandwidth** and show that the optimal bandwidth can be written as $h_r = \sigma C_K n^{-1/5}$, where the constant C_K depends only on the kernel function. For the Gaussian kernel, show that $C_K \approx 1.059$.

Hint: Start from equation (17.11). Compute $R(f'')$ explicitly under the assumption that f is the normal density, and substitute the result into the optimal bandwidth formula. You may use the following standard Gaussian integrals for n even:

$$\int_{-\infty}^{\infty} z^n e^{-z^2}\, dz = \frac{(n-1)!!}{2^{n/2}}\sqrt{\pi}$$

where $(n-1)!! = (n-1)(n-3)\cdots 3 \cdot 1$ is the double factorial.

Specifically: $\int_{-\infty}^{\infty} e^{-z^2}\, dz = \sqrt{\pi}$, $\int_{-\infty}^{\infty} z^2 e^{-z^2}\, dz = \frac{\sqrt{\pi}}{2}$, $\int_{-\infty}^{\infty} z^4 e^{-z^2}\, dz = \frac{3\sqrt{\pi}}{4}$.