Midterm Spring 2016

Q1. (25 pts.) Consider the simple regression $y_i = \beta x_i + \varepsilon_i$ where $\mathrm{E}[\varepsilon|x] = 0$ and $\mathrm{E}[\varepsilon^2|x] = \sigma^2$.

a. What is the minimum mean squared error linear estimator of β? [*Hint*: Let the estimator be ($\hat{\beta} = \boldsymbol{c}'\boldsymbol{y}$). Choose $\boldsymbol{c}$ to minimize $Var(\hat{\beta}) + (\mathrm{E}(\hat{\beta} - \beta))^2$. The answer is a function of the unknown parameters.]

b. For the estimator in part a, show that ratio of the mean squared error of $\hat{\beta}$ to that of the ordinary least squares estimator b is

$$\frac{MSE[\hat{\beta}]}{MSE[b]} = \frac{\tau^2}{1+\tau^2}, \quad \text{where } \tau^2 = \frac{\beta^2}{[\sigma^2/\boldsymbol{x}'\boldsymbol{x}]}.$$

Note that τ is the square of the population analog to the "t ratio" for testing the hypothesis that $\beta = 0$. How do you interpret the behavior of this ratio as $\tau \to \infty$?

Q2. (20 pts.) Let e_i be the ith residual in the ordinary least squares regression of $\boldsymbol{y}$ on $\boldsymbol{X}$ in the classical regression model, and let ε_i be the corresponding true disturbance. Prove that $\mathrm{plim}(e_i - \varepsilon_i) = 0$.

Q3. (15 pts.) For the simple regression model $y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim N[0, \sigma^2]$, prove that the sample mean is consistent and asymptotically normally distributed. Now consider the alternative estimator $\hat{\mu} = \sum_i w_i y_i$, $w_i = \frac{i}{n(n+1)/2} = \frac{i}{\sum_i i}$. Note that, $\sum_i w_i = 1$. Prove that this is a consistent estimator of μ and obtain its asymptotic variance. [*Hint*:$\sum_i i^2 = n(n+1)(2n+1)/6$.]

Q4. (20 pts.) Consider the model:

$$y = \beta x^* + \gamma d + \varepsilon, \quad x = x^* + u.$$

Assume that x^*, ε, and u are independent normally distributed with zeros means. Suppose that d is a random variable that takes the values one and zero with probabilities π and $1-\pi$ in the population and is independent of all other variables in the model. To put this formulation in context, the preceding model (and variants of it) have appeared in the literature on discrimination. We view y as "wage" variable, x^* as "qualifications," and x as some imperfect measure such as education. The dummy variable d is membership ($d=1$) or nonmembership ($d=0$) in some protected class. The hypothesis of discrimination turns on $\gamma < 0$ versus $\gamma \geq 0$.

a. What is the probability limit of c, the least squares estimator of γ, in the least squares regression of y on x and d? [*Hints*: The independence of x^* and d is important. Also, plim $\boldsymbol{d'd}/n = Var[d] + \mathrm{E}^2[d] = \pi(1-\pi) + \pi^2 = \pi$. This minor modification does not affect the model substantively, but it greatly simplifies the algebra.] Now suppose that x^* and d are not independent. In particular, suppose that $\mathrm{E}[x^*|d=1] = \mu^1$ and $\mathrm{E}[x^*|d=0] = \mu^0$. Repeat the derivation with this assumption.

b. Consider, instead, a regression of x on y and d. What is the probability limit of the coefficient on d in this regression? Assume that x^* and d are independent.

c. Suppose that x^* and d are not independent, but γ is, in fact, less than zero. Assuming that both preceding equations still hold, what is the estimated by $(\bar{y}|d=1)-(\bar{y}|d=0)$? What does this quantity estimate if γ does equal to zero?

Q5. (20 pts.) Find the asymptotic distribution of $\sqrt{n}(\hat{\sigma}^2 - \sigma^2)$ as $n \to \infty$.