Econometrics 522A
Spring, 2019

Mid-Term Exam

1. You are given the following empirical results from the regression model: Sample size is 1000.

$$\widehat{\log wage} = \underset{(0.002)}{0.542} - \underset{(0.007)}{0.309}\text{Female} + \underset{(0.018)}{0.010}\text{AA} + \underset{(0.018)}{0.118}\text{Ever Married}$$
$$+ \underset{(0.015)}{0.307}\text{Some College} + \underset{(0.014)}{0.211}\text{Aged 35-54} + \underset{(0.018)}{0.124}\text{Aged55+},$$

where numbers in parenthesis denote standard error of the estimates and "Female" and "AA" indicate dummy variables indicating females and African Americans, respectively, "Aged 55+" indicates the dummy variable denoting the group above age 55.

   (a) Explain why there is no dummy variable indicating the group age 34 and below.

   (b) Give an interpretation of the coefficient on the dummy variable "Aged 55+".

   (c) Give an interpretation of the coefficient on the dummy variable "Female".

   (d) Test the null hypothesis that there is no gender discrimination using the result against one-sided alternative. Be specific about the degrees of freedom.

   (e) Comment on the test in (d) from the perspective that the gender discrimination may happen in returns to education.

   (f) Does your test differ if you are told that the standard error in parenthesis are computed based on heteroskedasticity robust variance-covariance matrix estimator? Briefly explain your answer.

   (g) Construct a 95% confidence interval for the return to some college education.

   (h) In this regression model, there is no variable indicating a type of job for each worker. Explain sufficient conditions under which the omission does not lead to inconsistency of the OLS estimator.

   (i) Suppose you estimate the regression model using male dummy variable, instead of the female dummy variable, what is the coefficient estimate on the male dummy variable? Is there any other changes?

   (j) Explain how you will test if age affects wage allowing for heteroskedasticity. Are the pieces of information given here enough to compute the relevant test statistic? If not, explain what you need.

   (k) Suppose you think that the "ability" should be included in the regression model but that you do not have the variable capturing it. Explain how you will assess the potential bias in the return to education due to the omitted variable in this model.

   (l) Suppose you want to replicate the above result using a different data set but found that in the new data set, everyone who are aged 35 and above are also married at some point and vice versa. Namely, no one below 35 is married. Can you test the gender discrimination in this case? Explain your answer.

2. Please answer the following questions:

   (a) Give an example of a sequence of random variables which is $o_p(1)$. A sequence of numbers is not allowed.

   (b) Given an example of a sequence of random variables which is $O_P(1)$ but not $o_p(1)$. A sequence of numbers is not allowed.

   (c) Given an example of a sequence of random variables which is $O_P(1)$ but does not converge in distribution.

   (d) Give an example of a sequence of random variables which converges in distribution to a random variable but does not converge in probability.

(e) Show that if $\hat{v}_n$ converges in probability to v_0 which is bounded and strictly positive, and $\sqrt{n}(\hat{b}-\beta)$ converges in distribution to a normal random variable with mean zero and variance v_0, then $\sqrt{n}(\hat{b}-\beta)/\sqrt{\hat{v}_n}$ converges in distribution to the standard normal random variable.

(f) Show that a sequence of normal random variable with the mean between 0 and 5 and variance 1 is $O_P(1)$.

Problem 1 Midterm Sol'n's

(a) $d_{1i} + d_{2i} + d_{3i} = 1$ multicollinearity

$d_1 = 1 = \log wage = \Pi_0 + \Pi_1$

$d_2 = \ldots = \Pi_0 + \Pi_2$

$d_3 = \ldots = \Pi_0 + \Pi_3$

$\alpha = \beta_2 d_2 + \beta_3 d_3$

$\alpha = \Pi_0 + \Pi_1$

$\beta_2 = \Pi_2 - \Pi_1 \quad \beta_3 = \Pi_3 - \Pi_1$

(b) β_3 on Aged 55+ is effect of being that age on log(wage)

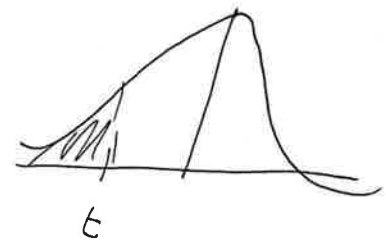
where $\beta_3 = \Pi_3 - \Pi_1$ being younger than 34, not affect of being younger than 55.

34 is reference group + we compare to reference group w/ dummies

(d) $H_0: \beta = 0$

$H_1: \beta_1 < 0$

$\frac{-0.309}{0.007} \sim t_{(1000-7)}$



note: there is no $\sqrt{1000-7}$

t test is justified now if we don't assume normality

$(X'X)^{-1} \rightarrow n E(x_i x_i')^{-1}$

$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \sigma^2 E(x_i x_i')^{-1})$

$(\hat{\beta} - \beta_0) \sim N(0, \sigma^2 (X'X)^{-1})$

ε norm

this cancels in this case

(e) Female · Some College
Female $\Big\}$ test jointly significant w/ F-test

$C' = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(f) Test will not differ.

(g) $0.307 \pm 1.96 \cdot 0.015$ No dividing by $\sqrt{n}$

(h) uncorrelated (as long not violates ~~[illegible]~~ conditional mean function)

(i) $\alpha + \beta d_i \quad d_i = 1$ female

$\Rightarrow \alpha + \beta(1 - d_i) = (\alpha + \beta) - \beta d_i$

(j) Would use F-test, but don't know var-cov matrix for

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \sim N\left(0, \begin{bmatrix} \hat{\sigma}_1^2 & cov \\ cov & \hat{\sigma}_2^2 \end{bmatrix}\right)$$

Don't know $Var(\hat{\beta}_1 + \hat{\beta}_2) = \underbrace{Var(\hat{\beta}_1)}_{\text{std error}} + \underbrace{Var(\hat{\beta}_2)}_{\text{std error}} + \underbrace{Cov(\hat{\beta}_1, \hat{\beta}_2)}_{\text{Not given in question here}}$

(k) $\tilde{\beta} = \hat{\beta} + \hat{\beta}_0 \hat{\pi}$ for var X in the regression

$\tilde{\beta}$: inconsistent estimates for x in the short regression

$\hat{\beta}$: consistent estimate for x in the long (full) regression.

$\hat{\pi}$ is estimate for X in the auxiliary regression of omitted variable

β_0 should be positive since ability should increase your wage in most cases.

$\hat{\beta}_0$: +

if x = some college :

$\hat{\beta}$: + $\hat{\pi}$: +

$$\underset{+}{\tilde{\beta}} = \underset{+}{\hat{\beta}} + \underset{+}{\hat{\beta}_0}\,\underset{+}{\hat{\pi}}$$ estimate will be upwards biased

(l) Aged 35-54 + Aged 55+ = evermarried

$\Rightarrow$ multicollinearity. Can only identify 2 of them

Its okay if you want to identify gender discrimination, since age does matter in that case.

Problem 2

(a) $X_n(\omega) = \begin{cases} \frac{1}{n} & \omega \in [0, \frac{1}{n}] \\ 0 & \omega \in [\frac{1}{n}, 1] \end{cases}$ $\qquad o_p(1)$

(b) $X_n(\omega) = \begin{cases} 1 + \frac{1}{n} & \omega \in [0, \frac{1}{n}] \\ 1 & \omega \in [\frac{1}{n}, 1] \end{cases}$ $\qquad O_p(1)$

$0 < X_n(\omega) < 2$

(c) $X_n \sim N(\mu_n, 1)$

$\mu_n \in [0, 5]$

(d) Converge in dist, not prob

$X_n = (-1)^n X_\infty$ $\qquad -N(0,1) \sim N(0,1)$ symmetric about 0

$X_\infty = N(0,1) \Rightarrow X_n \xrightarrow{d} X_\infty \sim N(0,1)$

$\Pr(|X_n - X_\infty| < \varepsilon) = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{odd} \end{cases}$

$X_n \not\xrightarrow{p} X_\infty$

(e)

$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, V_0)$ ①

$\hat{V}_0 \xrightarrow{p} V_0$ + CMT $\Rightarrow$ ② $\sqrt{\hat{V}_n} \xrightarrow{p} \sqrt{\hat{V}_0}$ $\qquad f(\hat{V}_n) = \sqrt{\hat{V}_n}$

(1) + (2) + Slutzsky $\dfrac{\sqrt{n}(\hat{\beta} - \beta)}{\sqrt{\hat{V}_0}} \xrightarrow{d} \dfrac{N(0, V_0)}{\sqrt{V_0}} = N(0,1)$

(f) $x_i \sim N(\mu_i, 1)$

$$X_i = \mu_i + (x_i - \mu_i) \sim N(0,1) = O_p(1)$$

$$\mu_i \in [0,5] \quad \mu_i := O_p(1)$$

$$O_p(1) + O_p(1) = O_p(1)$$