ECON 466 MIDTERM II ANSWER KEY

LIANG LIU

Question 1.

1. (i)

Table 1 Predicted logW for 10 years of education			
	Asian	Black	White
Male	1.4	1.05	1.25
Female	1.27	0.87	1.1
Difference	0.13	0.18	0.15

1. (ii) The reference group is White-male group.

1. (iii) For convenience,

$$\hat{logW} = \hat{\beta}_0 + \hat{\beta}_1 D_F + \hat{\beta}_2 D_B + \hat{\beta}_3 D_A + \hat{\beta}_4 D_F * D_A + \hat{\beta}_5 D_F * D_B + \hat{\beta}_6(educ - 10)$$

where $\hat{\beta}_0 = 1.25, \hat{\beta}_1 = -0.15, \hat{\beta}_2 = -0.2, \hat{\beta}_3 = 0.15, \hat{\beta}_4 = 0.02, \hat{\beta}_5 = -0.03, \hat{\beta}_6 = 0.06$

To make things clearer, I reconstruct the table above in $\hat{\beta}$ form

Table 2 Predicted logW for 10 years of education			
	Asian	Black	White
Male	$\hat{\beta}_0 + \hat{\beta}_3$	$\hat{\beta}_0 + \hat{\beta}_2$	$\hat{\beta}_0$
Female	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_4$	$\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_5$	$\hat{\beta}_0 + \hat{\beta}_1$
Difference	$\hat{\beta}_1 + \hat{\beta}_4$	$\hat{\beta}_1 + \hat{\beta}_5$	$\hat{\beta}_1$

Interpretation:

Ceteris paribus,

$\hat{\beta}_1 = -0.15$: on average the White-females are estimated to earn 15% less than White-males.

(Note: $\hat{\beta}_1$ is the difference between White-female and White-male from above table.)

$\hat{\beta}_2 = -0.2$: on average the Black-males are estimated to earn 20% less than White-males.

(Note: $\hat{\beta}_2$ is the difference between Black-male and White-male, similarly.)

$\hat{\beta}_3 = 0.15$: on average the Asian-males are estimated to earn 15% more than White-males.

(Note: $\hat{\beta}_3$ is the difference between Asian-male and White-male.)

$\hat{\beta}_4 = 0.02$: the average percentage difference between Asian-female and Asian-male is predicted to be 2 percentage points less than that between White-female and White-male.

(Note: $\hat{\beta}_4$ is the difference of the male-female differences across racial.)

$\hat{\beta}_5 = -0.03$: the average percentage difference between Black-female and Black-male will be 3 percentage points more than that between White-female and White-male.

$\hat{\beta}_6 = 0.06$: Having one more schooling year, the model predicts that people will earn 6% more.

1. (iv) The differences are shown in Table 1. Obviously they are different. Again through Table 2, we could see the differences of the differences across racial groups are $\hat{\beta}_4$ or $\hat{\beta}_5$, so the null hypothesis will be $H_0 : \beta_4 = \beta_5 = 0$ against $H_1 : H_0$ is not true. We will use F test, since there are two constraints. The regression in (1) is the unrestricted model. We need a restricted model, which is $\hat{logW} = \hat{\beta}_0 + \hat{\beta}_1 D_F + \hat{\beta}_2 D_B + \hat{\beta}_3 D_A + \hat{\beta}_6(educ - 10)$. Then we could calculate the F statistics. Comparing with the critical value $(F_{2,n-7})$, if $F > F_{2,n-7}$, we will reject the null.

1. (v) For better illustration, I use $\hat{\gamma}$s represent the unknown coefficients in model (2)

$$\hat{logW} = \hat{\gamma}_0 + \hat{\gamma}_1 Asianmale + \hat{\gamma}_2 Asianfemale + \hat{\gamma}_3 Blackmale$$
$$+\hat{\gamma}_4 Blackfemale + \hat{\gamma}_5 Whitemale + \hat{\gamma}_6(educ - 10)$$

Table 3 Predicted logW for 10 years of education			
	Asian	Black	White
Male	$\hat{\gamma}_0 + \hat{\gamma}_1$	$\hat{\gamma}_0 + \hat{\gamma}_3$	$\hat{\gamma}_0 + \hat{\gamma}_5$
Female	$\hat{\gamma}_0 + \hat{\gamma}_2$	$\hat{\gamma}_0 + \hat{\gamma}_4$	$\hat{\gamma}_0$

Comparing Table 1 and Table 3, we could get:

$$\hat{\gamma}_0 + \hat{\gamma}_1 = 1.4$$
$$\hat{\gamma}_0 + \hat{\gamma}_2 = 1.27$$
$$\hat{\gamma}_0 + \hat{\gamma}_3 = 1.05$$
$$\hat{\gamma}_0 + \hat{\gamma}_4 = 0.87$$
$$\hat{\gamma}_0 + \hat{\gamma}_5 = 1.25$$
$$\hat{\gamma}_0 = 1.1$$

so the value of $\hat{\gamma}$s are straightforward,

$$\begin{aligned} lo\hat{g}W = 1.1 + 0.3Asianmale + 0.17Asianfemale - 0.05Blackmale \\ -0.23Blackfemale + 0.15Whitemale + 0.06(educ - 10) \end{aligned}$$

1. (vi) We need only a gender dummy, $D_F = 1$, if female. The model will be:

$$logW = \delta_0 + \delta_1 educ + \delta_2 D_F + \delta_3 D_F * educ + error$$

To test that returns from education do not differ between the genders ($H_0 : \delta_3 = 0$ against $H_1 : \delta_3 \neq 0$), so we could use either t test or F test, we could get the t statistics from above regression directly or we could run a restricted regression excluding the interaction term ($\delta_3 D_F * educ$) and then compare with $F_{1,n-4}$. We decide whether reject the null as usual.

Question 2.

2. (i) I expect signs on the coefficients are positive.

β_1: Positive. Given education and age, good-looking ones often earn more.

β_2: Positive. Within some range, the person will earn more with more schooling years.

β_3: Positive. Within some range, the elder person has more experience, which will bring him/her higher wage.

2. (ii) We need to incorporate a gender dummy($D_F = 1$, if female). The model will be

$$log(wage) = \gamma_0 + \gamma_1 beauty + \gamma_2 educ + \gamma_3(age - 20) + \gamma_4 D_F * beauty + error$$

We will test $H_0 : \gamma_4 = 0$ against $H_1 : \gamma_4 \neq 0$. I am going to use t test. Comparing with t_{n-5} given certain α, if $t > t_{n-5}$, we will reject the null, which implies the "beauty-effect" does exist more on women.

2. (iii) We could use either Breusch-Pagan or White test. I am using Breusch-Pagan test here.

Step 1: Estimate (3) by OLS and obtain the residuals $\hat{u}_i$, $i = 1, ..., n$, and obtain $\hat{u}_i^2$.

Step 2: Regress $\hat{u}_i^2$ on the explanatory variables as

$$\hat{u}_i^2 = \delta_0 + \delta_1 beauty + \delta_2 educ + \delta_3(age - 20) + error$$

obtaining the $R^2_{\hat{u}_i}$.

Step 3: Get F (or LM) statistics using $R^2_{\hat{u}_i}$, in which the restricted model is no regression. $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$ against $H_1 : H_0$ is not true.

If the F (or LM) statistics obtained above exceeds the critical value ($F_{3,n-4}$)at the chosen level of significance, the conclusion is that there is heteroscedasticity. (Or if the p-value of the statistics is sufficiently small, or below the chosen significance level, then we reject the null hypothesis of homoscedasticity.)

Correcting for the heteroscedasticity with unknown $var(u_i)$, given $var(u|x) = \sigma^2 exp(\gamma + \delta D_m) \equiv \sigma^2 h(D_M)$, given $h(D_M) = exp(\gamma + \delta D_M)$ with unknown parameters. We will use Feasible GLS here.

Step 1: Estimate (3) by OLS and obtain the residuals $\hat{u}_i$, $i = 1, ..., n$, and obtain $\hat{u}_i^2$.

Step 2: Take natural log of $\hat{u}_i^2$.

Step 3: Estimate the following model by OLS and get the fitted value of $\hat{g}_i \equiv \hat{log(\hat{u}_i^2)}$

$$log(\hat{u_i^2}) = \alpha_0 + \gamma D_M + error$$

Step 4: Exponentiate the fitted values: $\hat{h}_i = exp(\hat{g}_i)$

Step 5: Estimate (3) by WLS using weights $1/\hat{h}_i$, specifically:
Divided both side of equation (3) by $\sqrt{\hat{h}_i}$ to get

$$\frac{log(wage)}{\sqrt{\hat{h}_i}} = \frac{\beta_0}{\sqrt{\hat{h}_i}} + \beta_1 \frac{beauty}{\sqrt{\hat{h}_i}} + \beta_2 \frac{educ}{\sqrt{\hat{h}_i}} + \beta_3 \frac{(age-20)}{\sqrt{\hat{h}_i}} + \frac{u_i}{\sqrt{\hat{h}_i}}$$

So we will run a new regression, which regress $\frac{log(wage)}{\sqrt{\hat{h}_i}}$ on $\frac{1}{\sqrt{\hat{h}_i}}, \frac{beauty}{\sqrt{\hat{h}_i}}, \frac{educ}{\sqrt{\hat{h}_i}}$, and $\frac{(age-20)}{\sqrt{\hat{h}_i}}$ without intercept.

Question 3.

3. (a) Ignoring the fact that two sequences are trending in the same or opposite directions can lead us to falsely conclude that changes in one variable are actually caused by changes another variables. In many cases, two time series processes appear to be correlated only because they are both trending over time for reasons related to other unobserved factors. We could avoid the problem by detrending. Given linear trend, the detrending is simply adding a t variable into the original regression, which will capture the trends of all involved trending sequences.

3. (b) It depends. Define the measurement error as $e_t = Z_t - Z_t^*$, where Z_t^* is the true value and Z_t is the one with measurement error. The effect of measurement error on OLS estimates depends on the assumptions about the correlation between e_t and Z_t. If $Cov(Z_t, e_t) = 0$, OLS $\hat{\beta}_1$ is consistent. If the $Cov(Z_t^*, e_t) = 0, Cov(Z_t, e_t) = E(Z_t e_t) = E(Z_t^* e_t) + E(e_t^2) = 0 + \sigma_{e_t}^2 = \sigma_{e_t}^2$ therefore the OLS $\hat{\beta}_1$ will be biased, then inconsistent.

3. (c) In the presence of serial correlation, the usual OLS standard error will be invalid. Therefore the usual t, F and LM statistics will be invalid also.

We could use serial correlation-robust standard error to solve the problem without correcting for autocorrelation. But it is not as popular as heteroscedasticity-robust standard error.

3. (d) Testing for AR(1) serial correlation:

Step 1: Run the OLS regression: $Y_t = \beta_0 + \beta_1 Z_t + u_t$ and get the OLS residuals, $\hat{u}_t$, for all t=1,...n.

Step 2: Run the regression: $\hat{u}_t = \rho u_{t-1} + e_t$, obtaining the coefficient $\hat{\rho}$ and its t statistics.

Step 3: Use $\hat{t}_{\hat{\rho}}$ to test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$ in the usual way.

If AR(2),

Step 1: The same as above.

Step 2: We will change the model in Step 2 above to:

$\hat{u}_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t$, obtaining the F statistics, and restricted model is no regression.

Step 3: Use F statistics to test $H_0 : \rho_1 = 0, \rho_2 = 0$ against $H_1 : H_0$ is not true in the usual way.

3. (e) We use Feasible GLS to estimate the model:

Step 1: Step 1: Run the OLS regression: $Y_t = \beta_0 + \beta_1 Z_t + u_t$ and get the OLS residuals, $\hat{u}_t$, for all t=1,...n.

Step 2: Run the regression: $\hat{u}_t = \rho u_{t-1} + e_t$ and obtain $\hat{\rho}$.

Step 3: Manipulate the data set as:

$$Y_{t-1} = \beta_0 + \beta_1 Z_{t-1} + u_{t-1}$$
$$Y_t = \beta_0 + \beta_1 Z_t + u_t$$

Multiplying the first equation above by ρ and subtracting it from the second equation, we get

$Y_t - \rho Y_{t-1} = (1 - \rho)\beta_0 + \beta_1(Z_t - \rho Z_{t-1}) + e_t, t \geq 2$, where we have used the fact

$e_t = u_t - \rho u_{t-1}$. We use the manipulated data (quasi-differenced data) to run OLS to estimate the βs. The above procedure is called Cochrane-Orcutt(CO) estimation. (You could use Prais-Winsten(PW) estimation also.)

3. (f) Given MA(1) process, $u_t = \varepsilon_t - \lambda\varepsilon_{t-1}$, where $\varepsilon_t \sim i.id.(0, \sigma_\varepsilon^2)$

$E(u_t) = E(\varepsilon_t - \lambda\varepsilon_{t-1}) = 0$

$u_{t+1} = \varepsilon_{t+1} - \lambda\varepsilon_t$

$u_{t+2} = \varepsilon_{t+2} - \lambda\varepsilon_{t+1}$

$$\begin{aligned} cov(u_t, u_{t+1}) &= E[(u_t - 0)(u_{t+1} - 0)] \\ &= E[(\varepsilon_t - \lambda\varepsilon_{t-1})(\varepsilon_{t+1} - \lambda\varepsilon_t)] \\ &= E(\varepsilon_t\varepsilon_{t+1} - \lambda\varepsilon_{t-1}\varepsilon_{t+1} - \lambda\varepsilon_t^2 + \lambda^2\varepsilon_{t-1}\varepsilon_t) \\ &= 0 - 0 - \lambda\sigma_\varepsilon^2 + 0 \\ &= -\lambda\sigma_\varepsilon^2 \end{aligned}$$

$$\begin{aligned} cov(u_t, u_{t+2}) &= E[(u_t - 0)(u_{t+2} - 0)] \\ &= E[(\varepsilon_t - \lambda\varepsilon_{t-1})(\varepsilon_{t+2} - \lambda\varepsilon_{t+1})] \\ &= E(\varepsilon_t\varepsilon_{t+2} - \lambda\varepsilon_{t-1}\varepsilon_{t+2} - \lambda\varepsilon_t\varepsilon_{t+1} + \lambda^2\varepsilon_{t-1}\varepsilon_{t+1}) \\ &= 0 - 0 - 0 + 0 \\ &= 0 \end{aligned}$$

Department of Economics