Course: ECON 613

Name: Yuqi Zhou

Date: 02/04/2022

Project: Assignment 2

Exercise 1 OLS estimate

• Calculate the correlation between Y and X.

The correlation between Y and X is 0.143492.

• Calculate the coefficients on this regression.

The coefficient is 230.9923.

• Calculate the standard errors of β

The standard error is 509.2344.

```
Residuals:
   Min     1Q Median     3Q     Max
-37130 -10653  -2196    6030 361786

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 14141.18     645.23   21.92   <2e-16 ***
age           230.99      14.88   15.53   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19480 on 11467 degrees of freedom
Multiple R-squared:  0.02059,   Adjusted R-squared:  0.0205
F-statistic: 241.1 on 1 and 11467 DF,  p-value: < 2.2e-16
```
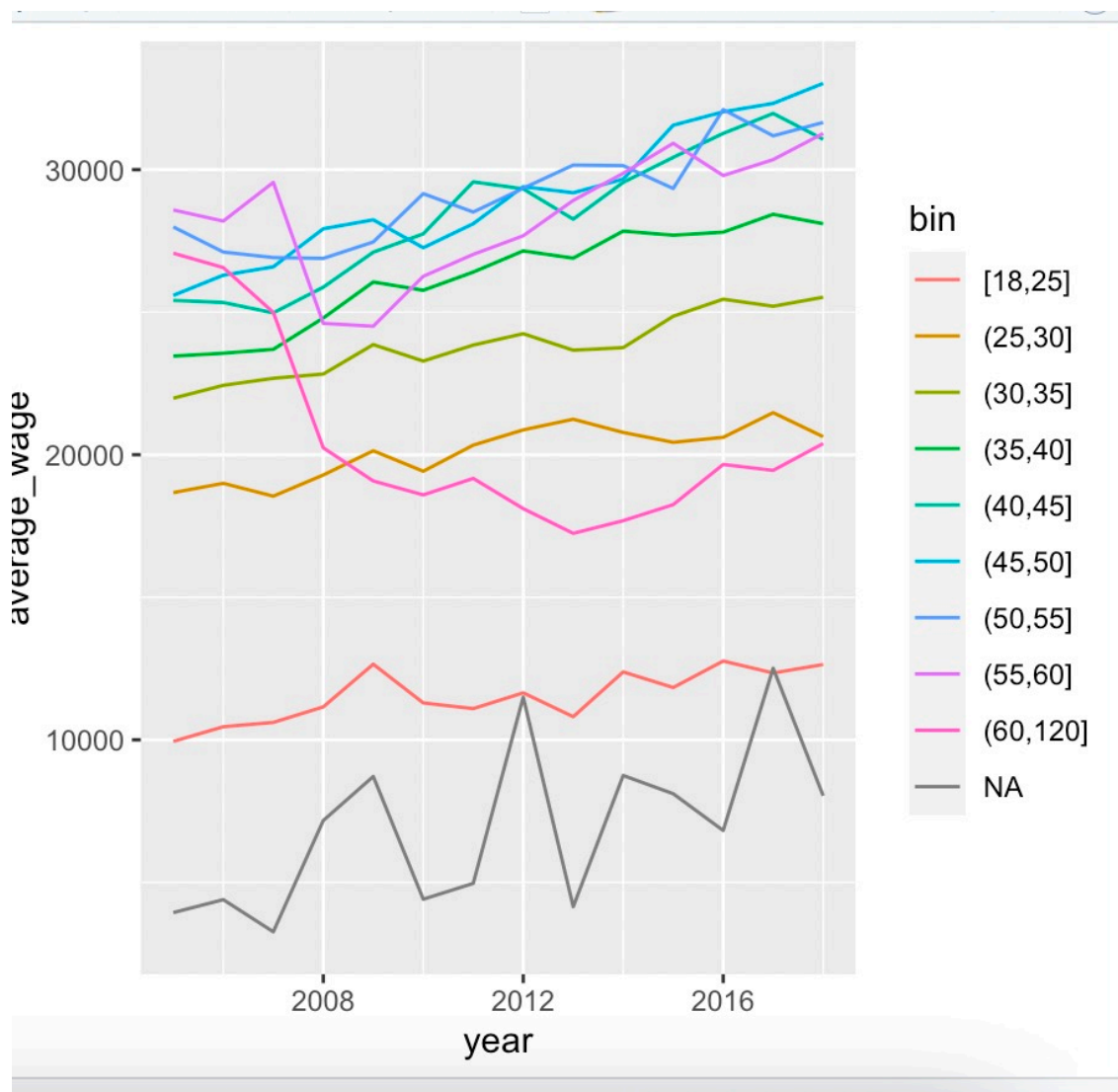
The two strategies give similar results. The accuracy increases when the number

of bootstrap increases.

Exercise 2 Detrend Data

• Create a categorical variable ag, which bins the age variables into the following

  groups: "18-25", "26- 30", "31-35", "36-40","41-45", "46-50","51-55", "56-60", and

  "60+".

Presented in R programming.

• Plot the wage of each age group across years. Is there a trend?

The wage increases when people get older, and the average wage level increases through time.

- Consider Yit = βXit + γt + eit. After including a time fixed effect, how do the estimated coefficients change?

After including a time fixed effect, the estimated coefficient of age becomes larger.

Exercise 3 Numerical Optimization

- Exclude all individuals who are inactive.

Presented in R programming

- Write a function that returns the likelihood of the probit of being employed. You might want to write Xβ first. Then, calculate F(Xβ) and the log likelihood. Remember, for the probit model, F(x) is the standard normal distribution function.

Presented in R programming

- Optimize the model and interpret the coefficients. You can use pre-programmed optimization packages.

	V1	V2	V3
55	1.087762	0.01320048	2079.099

From the output, we can see that the age coefficient is positive, implying that age

has a positive effect for labors to be employed.

• Can you estimate the same model including wages as a determinant of labor

   market participation? Explain

No, I can't estimate the same model when including wages as a determinant.

Because unemployed people have no wage, and this would influence our

outcome.

Exercise 4 Discrete choice

• Exclude all individuals who are inactive.

Presented in R programming

• Write and optimize the probit, logit, and the linear probability models.

   Remember, for the logit model, F(x) is the logistic function exp(x) (1+exp(x))

Presented in R programming

• Interpret and compare the estimated coefficients. How significant are they?

```
> output5[which.max(output5$V5 == min(output5$V5)), ]
         V1        V2        V3        V4        V5
3 -4.151691 8.271051 6.096356 3.766964 79563.63
```


```
> output6[which.max(output6$V5 == min(output6$V5)), ]
         V1        V2        V3        V4        V5
2 -7.712691 0.3578611 8.198586 6.674762 79563.63
```

```
> output7[which.max(output7$V5 == min(output7$V5)), ]
        V1        V2        V3       V4       V5
1 9.786648 6.747987 9.446685 1.54806 79563.63
>
```

In all methods, age appeared to have positive and significant effects.

Exercise 5 Marginal Effects

• Compute the marginal effect of the previous probit and logit models.

• Construct the standard errors of the marginal effects.

```
> m_probit(fun <- empstat2 ~ age2 + year, data <- datind_05_to_15)
             Marginal Effect           SE
(Intercept)   -8.9274778376 2.237966e-01
age2           0.0003351427 1.342062e-07
year           0.0044732472 1.113388e-04
>
```

```
> m_logit(formula = empstat2 ~ age2 + year, data = datind_05_to_15)
             Marginal Effect           SE
(Intercept)   -1.055478e+01 2.109887e-01
age2           3.668174e-04 9.415952e-08
year           5.276457e-03 1.049657e-04
~
```