Econ613 HW2

Yonghan Shi

1/30/2022

Exercise 1 OLS estimate

Consider the following model

$$Y = X\beta + \epsilon$$

where X is the age of individuals plus intercept, and Y is the wage.

- Calculate the correlation between Y and X.

The correlation between Y and X is 0.1398806.

- Calculate the coefficients on this regression.

intercept	age
14362.56	226.3447

- Calculate the standard errors of β

  – Using the standard formulas of the OLS.

intercept	age
650.70041069506	14.9844872003543

  – Using bootstrap with 49 and 499 replications respectively. Comment on the difference between the two strategies.

	intercept standard errors	age standard error
49 bootstrap	596.6602	16.03316
499 bootstrap	590.9932	15.48878

With more replications, the standard error will be smaller.

Exercise 2 Detrend Data

Consider the same application as exercise 1 but using a pooled version of individual data from 2005 to 2018.

- 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+".

- Plot the wage of each age group across years. Is there a trend?
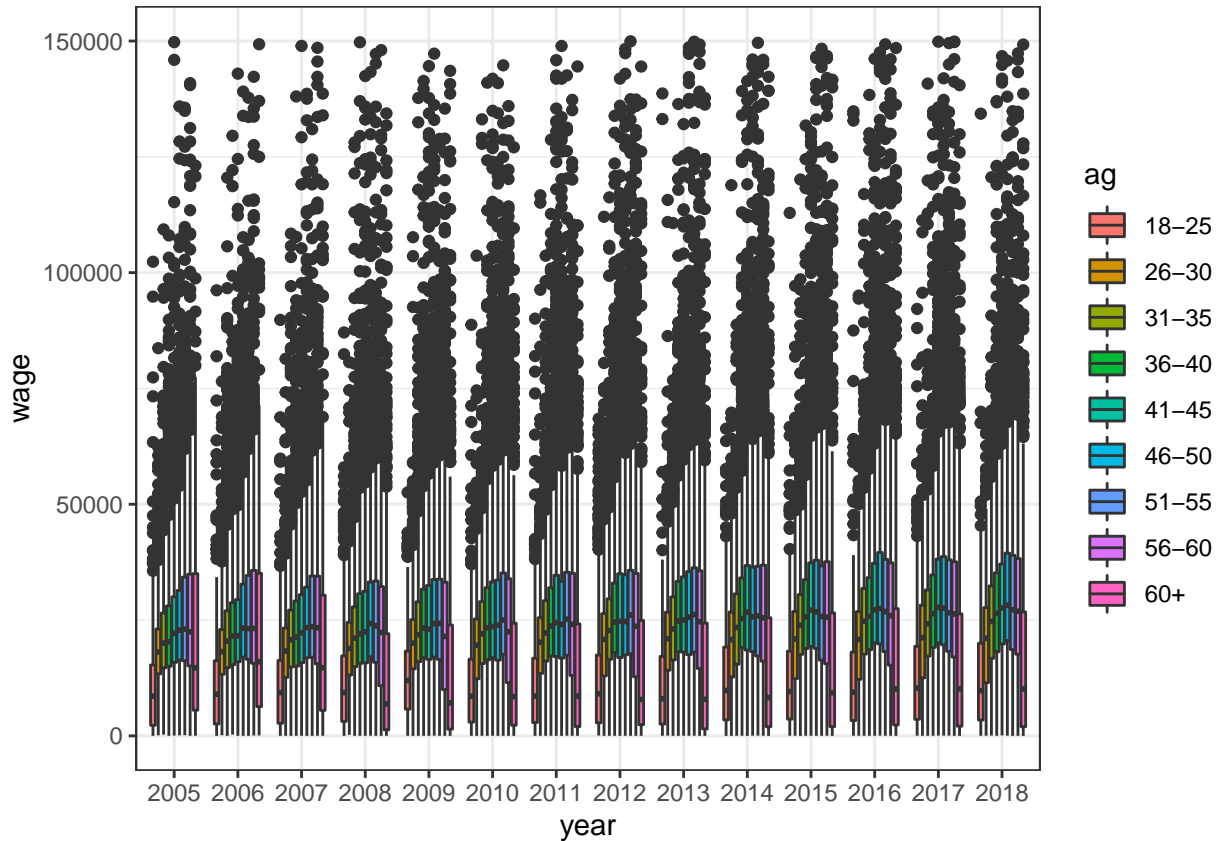
In general the wage peak of age group is moving from the elder to the middle ages. Younger people earn more during the period.

- Consider $Y_{it} = \beta X_{it} + \gamma_t + e_{it}$. After including a time fixed effect, how do the estimated coefficients change?

	intercept	age
Original model	11648.55	306.1192
Time fixed effect added	10625.07	297.8165

The coefficients of both age variable and intercept become smaller.

Exercise 3 Numerical Optimization

We are interested in the effect of age on labor market participation. We consider this problem using the data from 2007. Consider a probit model.

- Exclude all individuals who are inactive.

- Write a function that returns the likelihood of the probit of being employed.

You might want to write $X\beta$ first. Then, calculate $F(X\beta)$ and the log likelihood. Remember, for the probit model, $F(x)$ is the standard normal distribution function.

```
> X <- cbind(1, data3$age)
> Y <- data3$empstat
>
> #negative log-likelihood function
> probit.like <- function (beta) {
+     Xb <- X %*% beta
+     p <- pnorm(Xb)
```

```
+     -sum((1 - Y) * log(1 - p) + Y * log(p))
+ }
>
> #gradient function
> probit.grad <- function (beta) {
+     Xb <- X %*% beta
+     p <- pnorm(Xb)
+     u <- dnorm(Xb) * (Y - p) / (p * (1 - p))
+     -crossprod(X, u)
+ }
```

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

intercept	age
1.04524730990415	0.0068998883746946

The coefficients mean that with age becomes older, the probability of getting employed is higher.

- Can you estimate the same model including wages as a determinant of labor market participation? Explain.

I can't. As wage is highly correlated with probability of getting employed. The people who receive no wage would be those who are not employed.

Exercise 4 Discrete choice

We are interested in the effect of age on labor market participation. Use the pooled version of the data from 2005 to 2015. Additional controls include time-fixed effects.

- Exclude all individuals who are inactive.

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

Remember, for the logit model, $F(x)$ is the logistic function $\dfrac{exp(x)}{(1 + exp(x))}$

```
> X1 <- cbind(1, data4$age, model.matrix(~year -1, data4)[,2:11] )
> Y1 <- data4$empstat
>
> # probit model
>
> probit.li <- function (beta) {
+     est <- X1 %*% beta
+     p <- pnorm(est)
+     -sum((1 - Y1) * log(1 - p) + Y1 * log(p))
+ }
>
> probit.gr <- function (beta) {
+     est <- X1 %*% beta
+     p <- pnorm(est)
+     u <- dnorm(est) * (Y1 - p) / (p * (1 - p))
+     -crossprod(X1, u)
+ }
>
> b1 <- c(runif(1,0,1), runif(11,-0.1,0.1))
>
> fit1 <- optim(b1, probit.li, gr=probit.gr, method = "BFGS", hessian = TRUE)
>
> # logit model
```

```
> logit <- glm(data4$empstat ~ data4$age + data4$year, family = binomial(link = "logit") )
>
> logit.li <- function (beta) {
+    est <- X1 %*% beta
+    p <- 1/(1+exp(-est))
+    -sum((1 - Y1) * log(1 - p) + Y1 * log(p))
+ }
>
> logit.gr <- function (beta) {
+    est <- X1 %*% beta
+    p <- 1/(1+exp(-est))
+    u <- Y1 - p
+    -crossprod(X1, u)
+ }
>
> b2 <- c(runif(1,0,2), runif(11,-0.1,0.3))
>
> fit2 <- optim(b2, logit.li, gr=logit.gr, method = "BFGS", hessian = TRUE)
>
> # linear probability model
> olspar <- solve(t(X1) %*% X1) %*% (t(X1) %*% Y1)
```

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

	intercept	age
probit	0.7515654	0.0122596
logit	1.1230160	0.0252452
linear	0.7977484	0.0023359

The coefficients mean that with age becomes older, the probability of getting employed is higher. The coefficient of logit model is bigger than that of probit model, this is due to the difference $F(x)$ function of the two models.

Based on the Z-ratio of probit model and logit model we have in Excercise 5, we are more than 99% confident to reject the null hypothesis. They are significant.

Exercise 5 Marginal Effects

- Compute the marginal effect of the previous probit and logit models. • Construct the standard errors of the marginal effects. Hint: Boostrap may be the easiest way.

	marginal.effect	standard.error	z.ratio
probit	0.00218	0.000084	25.889103
logit	0.002319	0.000089	26.004959