Econometrics HW3

陈子睿 15220212202842

November 12, 2023

1 EX 6.2

(a) From the regression results in column (1), we can see that workers with college degrees can earn \$8.31/hour more than the workers with only high school degrees on average.

(b)Men earn \$3.85/hour more than women on average.

2 EX 6.3

(a) We can see that a worker earns \$0.51/hour more for each year he ages on average.

(b) Sally's earning prediction is given by

$$1.87 + 8.32 \times 1 - 3.81 \times 1 + 0.51 \times 29 = 21.17 \text{ dollars per hour.}$$

Betsy's earning prediction is given by

$$1.87 + 8.32 \times 1 - 3.81 \times 1 + 0.51 \times 34 = 27.32 \text{ dollars per hour.}$$

The difference is \$2.55/hour.

3 EX 6.4

(a) Workers in the Northeast earn \$0.18 more per hour than workers in the West, on average, controlling for other variables in the regression. Workers in the Midwest earn \$1.23 less per hour than workers in the West, on average, controlling for other variables in the regression. Workers in the South earn \$0.43 less than workers in the West, controlling for other variables in the regression.

(b) The regressor West is omitted to avoid perfect multicollinearity. If West is included, then the intercept can be written as a perfect linear function of the four regional regressors.

(c) The expected difference in earnings between Juanita and Jennifer is

-0.43 - (-1.23) = \$0.80/hour.

4 EX 6.11

Considering the regression model

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \text{ i= 1,...,n.}$$

(a) The LS function that minimized by OLS is

$$\sum_{i=1}^{n}(Y_i - b_1 X_{1i} - b_2 X_{2i})^2$$

(b)

$$\begin{cases} \frac{\partial}{\partial b_1}\sum_{i=1}^{n}(Y_i - b_1 X_{1i} - b_2 X_{2i})^2 = -2\sum_{i=1}^{n}(Y_i - b_1 X_{1i} - b_2 X_{2i})X_{1i} \\ \frac{\partial}{\partial b_2}\sum_{i=1}^{n}(Y_i - b_1 X_{1i} - b_2 X_{2i})^2 = -2\sum_{i=1}^{n}(Y_i - b_1 X_{1i} - b_2 X_{2i})X_{2i} \end{cases}$$

(c) From (b), we can see that $\hat{\beta}_1$ satisfies

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i})X_{1i} = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{1i}Y_i - \hat{\beta}_2 \sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{1i}^2}$$

Since we have $\sum_{i=1}^{n} X_{1i}X_{2i} = 0$, therefore

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{1i}Y_i}{\sum_{i=1}^{n} X_{1i}^2}$$

(d) Similar to (c), we have

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_{2i}Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{2i}^2}$$

Then we substituting the above equation into the expression for $\hat{\beta}_1$ in (c) and have

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{1i}Y_i - \frac{\sum_{i=1}^{n} X_{2i}Y_i - \hat{\beta}_1 \sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{2i}^2} \sum_{i=1}^{n} X_{1i}X_{2i}}{\sum_{i=1}^{n} X_{1i}^2}$$

Then we have

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_{2i}^2 \sum_{i=1}^{n} X_{1i}Y_i - \sum_{i=1}^{n} X_{1i}X_{2i} \sum_{i=1}^{n} X_{2i}Y_i}{\sum_{i=1}^{n} X_{1i}^2 \sum_{i=1}^{n} X_{2i}^2 - \left(\sum_{i=1}^{n} X_{1i}X_{2i}\right)^2}$$

(e) The least squares objective function is $\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2$ and the partial derivative with respect tob_0 is

$$\frac{\partial}{\partial b_0}\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2 = -2\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})$$

Solving for $\hat{\beta}_0$ yields that

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

(f) Substituting $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$ into the LS objective function yields that,

$$\sum_{i=1}^{n}(Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2 = \sum_{i=1}^{n}[(Y_i - \bar{Y}) - b_1(X_{1i} - \bar{X}_1) - b_2(X_{2i} - \bar{X}_2)]^2$$

which is identical to the least squares objective function in part (a), except that all variables have been replaced with deviations from sample means. The result then follows as in (c).

Notice that the estimator for β_1 is identical to the OLS estimator from the regression of Y onto X_1, omitting X_2. Said differently, when $\sum_{i=1}^{n}(X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$, the estimated coefficient on X_1 in the OLS regression of Y onto both X_1 and X_2 is the same as estimated coefficient in the OLS regression of Y onto X_1, omitting X_2.

5 EX ADDITIONAL

Firstly, income affects longevity, the rich may get better health care and live longer. Secondly, income is correlated with single, maybe it's easier for the rich to get married. Suppose the simple linear regression model is $Y_i = \beta_0 + \beta_1 X_i + u_i$, then the OLS estimator

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum X_i Y_i - \bar{X}\bar{Y}}{\frac{1}{n}\sum X_i^2 - \bar{X}^2} = \beta_1 + \frac{\frac{1}{n}\sum u_i(X_i - \bar{X})}{\frac{1}{n}\sum X_i^2 - \bar{X}^2} \xrightarrow{p} \beta_1 + \frac{Cov(X_i, u_i)}{Var(X_i)}$$

Since single is negatively correlated with longevity, i.e. $\beta_1 < 0$. The bias yield by omitted variable income is also negative, i.e. $Cov(X_i, u_i) < 0$, then $\hat{\beta}_1 < \beta_1$. Thus the OLS estimator $\hat{\beta}_1$ implies a larger effect than the true effect β_1.