1) In this problem we use hprice dataset. You can find this dataset and other datasets of "*Econometric Analysis of Cross Section and Panel Data*" here[1].

Use the data in hprice to estimate the model:

$$price = \beta_0 + \beta_1 area + \beta_2 rooms + u,$$

a) Write out the results in equation form.

b) What is the estimated increase in price for a house with one more bedroom, holding area constant?

c) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in part (b).

d) What percentage of the variation in price is explained by square footage and number of rooms?

e) Suppose a house in the sample has area = 2,438 and rooms = 4. Find the predicted selling price for this house from the OLS regression line.

f) The actual selling price of the first house in the sample was $60,000. Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

---

2) Use the data in "attend" to answer this question. You can find dataset similar to problem 1.

a) To determine the effects of attending lecture on final exam performance, estimate a model relating *stndfnl* (the standardized final exam score) to *atndrte* (the percent of lectures attended). Include the binary variables *frosh* and *soph* as explanatory variables. Interpret the coefficient on *atndrte*, and discuss its significance.

b) How confident are you that the OLS estimates from part a are estimating the causal effect of attendence? Explain.

c) As proxy variables for student ability, add to the regression *priGPA* (prior cumulative GPA) and *ACT* (achievement test score). Now what is the effect of *atndrte*? Discuss how the effect differs from that in part a.

d) What happens to the significance of the dummy variables in part c as compared with part a? Explain.

[1] In Stata you can use this command:
use http://www.stata.com/data/jwooldridge/eacsap/hprice

e) Add the squares of *priGPA* and *ACT* to the equation. What happens to the coefficient on *atndrte*? Are the quadratics jointly significant?

f) To test for a nonlinear effect of *atndrte*, add its square to the equation from part e. What do you conclude?

---

3) We want to regress y_i on x_{1i} and x_{2i} : $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$
We have 2 methods:

Method one) estimate it as a regression with 2 independent variables: $\hat{y}_i = \hat{\alpha} + \widehat{\beta_1} x_{1i} + \widehat{\beta_2} x_{2i}$

Method two) estimate $x_{1_i} = \gamma + \beta_{12} x_{2i} + x^*_{1i}$ and calculate estimate of residual $(\hat{x}^*_{1i})$
Then estimate $y_i = \alpha' + \beta_1{}' \hat{x}^*_{1i} + \beta'_2 x_{2i} + e_i$

a) show $\widehat{\beta'_1} = \widehat{\beta_1}$
b) if $y_i = \alpha'' + \beta''_2 x_{2i} + v_i$ show $\widehat{\beta'_2} = \widehat{\beta_2}''$
c) show $E\left(\widehat{\beta'_2}\right) \neq \beta_2$ and calculate biasedness of $\widehat{\beta'_2}$ for β_2.

PRE $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$ (3) .۳

سروش زاهدی
۴۰۲۲۱۱۰۲۴

$$\beta_1 = \frac{\partial E(y_i \mid x_{1i}, x_{2i})}{\partial x_{1i}}$$

(1) OLS $\longrightarrow$ $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$

$\longrightarrow$ LS residual

(2) $x_{1i} = \gamma + \beta_{12} x_{2i} + x_{1i}^*$ (1) $\rightarrow$ then regress the estimate of the residual $\hat{x}_{1i}^*$ on y_i

$y_i = \alpha' + \beta_1' \hat{x}_{1i}^* + e_i$ (2)

a) $\hat{\beta}_1 = \hat{\beta}_1'$

$$\hat{\beta}_1' = \frac{\sum(\hat{x}_{1i}^* - \bar{\hat{x}}^*)(y_i - \bar{y})}{\sum(\hat{x}_{1i}^* - \bar{\hat{x}}^*)^2}$$

From the F.O.C for the OLS of (1) we have:

$$\sum \hat{x}_{1i}^* = 0 \implies \bar{\hat{x}}^* = 0$$

$$\beta_1' = \frac{\sum \hat{x}_{1i}^*(y_i - \bar{y})}{\sum \hat{x}_{1i}^{*2}} = \frac{\sum \hat{x}_{1i}^* y_i}{\sum \hat{x}_{1i}^{*2}}$$ $\longrightarrow$ we have to show that $\hat{\beta}_1$ equals this

From the F.O.C of (3) we have:

$\sum \hat{u}_i x_{1i} = 0$, we know that $x_{1i} = \hat{x}_{1i} + \hat{x}_{1i}^*$

$$\sum \hat{u}_i x_{1i} = \sum \hat{u}_i (\hat{x}_{1i} + \hat{x}^*_{1i}) = 0$$

we know $\hat{x}_{1i} = \hat{\gamma} + \hat{\beta}_{12} x_{2i}$

$$\Rightarrow \sum \hat{u}_i \hat{x}_{1i} = \sum \hat{u}_i (\hat{\gamma} + \hat{\beta}_{12} x_{2i}) = \hat{\gamma} \underbrace{\sum \hat{u}_i}_{0} + \hat{\beta}_{12} \underbrace{\sum \hat{u}_i x_{2i}}_{0}$$

$$\Rightarrow \boxed{\sum \hat{u}_i \hat{x}_{1i} = 0}$$

from F.O.C OLS

$$\Longrightarrow \boxed{\sum \hat{u}_i \hat{x}^*_{1i} = 0}$$

now we know $\hat{u}_i = y_i - \hat{y}_i \Rightarrow \sum (y_i - \hat{y}_i) \hat{x}^*_{1i} = 0$

$$\rightarrow \sum (y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) \hat{x}^*_{1i} = 0$$

$$= \sum y_i \hat{x}^*_{1i} - \hat{\alpha} \underbrace{\sum \hat{x}^*_{1i}}_{0} - \hat{\beta}_1 \sum x_{1i} \hat{x}^*_{1i} - \hat{\beta}_2 \underbrace{\sum x_{2i} \hat{x}^*_{1i}}_{0} = 0$$

from ①'s F.O.C's we have

$$\Rightarrow \sum y_i \hat{x}^*_{1i} - \hat{\beta}_1 \sum x_{1i} \hat{x}^*_{i1} = 0$$

$$= \sum y_i \hat{x}^*_{1i} - \hat{\beta}_1 \sum (\hat{x}_{1i} + \hat{x}^*_{i1}) \hat{x}^*_{i1} = 0$$

$$= \sum y_i \hat{x}^*_{1i} - \hat{\beta}_1 \underbrace{\sum \hat{x}_{1i} \hat{x}^*_{1i}}_{0} - \hat{\beta}_1 \sum \hat{x}^{*2}_{i1}$$

$\rightarrow$ Sample Correlation between residuals and the predicted values of a regression is always zero.

$$\sum y_i \hat{x}_{i1}^{*} - \hat{\beta}_1 \sum \hat{x}_{i1}^{*2} = 0$$

$$\Rightarrow \boxed{\hat{\beta}_1 = \frac{\sum y_i \hat{x}_{i1}^{*}}{\sum \hat{x}_{i1}^{*2}} = \hat{\beta}_1'}$$

b) $\hat{\beta}_2' = \hat{\beta}_2''$

① OLS $\longrightarrow$ $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i}$

$\longrightarrow$ LS residual

② $x_{1i} = \gamma + \beta_{12} x_{2i} + x_{1i}^{*}$ ① $\rightarrow$ then regress the estimate of the residual $\hat{x}_{1i}^{*}$ on y_i

$y_i = \alpha' + \beta_1' \hat{x}_{1i}^{*} + \beta_2' x_{2i} + e_i$

③ $y_i = \alpha'' + \beta_2'' x_{2i} + v_i$ ③

from ① we know that $\hat{x}_{1i}^{*}$ and x_{2i} are uncorrelated

$\sum \hat{x}_{1i}^{*} x_{2i} = 0$

from the F.O.C for OLS of ③:

$\sum \hat{u}_i x_{2i} = 0 = \sum (y_i - \hat{y}_i) x_{2i}$

$$= \sum (y_i - \hat{\alpha}' - \hat{\beta}_1' \hat{x}^*_{1i} - \hat{\beta}_2' x_{2i}) x_{2i} = 0$$

$$= \sum y_i x_{2i} - \hat{\alpha}' \sum x_{2i} - \hat{\beta}_1 \underbrace{\sum \hat{x}^*_{1i} x_{2i}}_{0} - \hat{\beta}_2' \sum x_{2i}^2 = 0$$

$$\Rightarrow \sum y_i x_{2i} - \hat{\alpha}' \sum x_{2i} - \hat{\beta}_2' \sum x_{2i}^2 = 0$$

$$\Rightarrow \hat{\beta}_2' = \frac{\sum y_i x_{2i} - \hat{\alpha}' \sum x_{2i}}{\sum x_{2i}^2}$$

به طور مشابه داریم : $\sum v_i x_{2i} = 0$

$$= \sum (y_i - \hat{\alpha}'' - \hat{\beta}_2'' x_{2i}) x_{2i} = 0$$

$$= \sum y_i x_{2i} - \hat{\alpha}'' \sum x_{2i} - \hat{\beta}_2'' \sum x_{2i}^2 = 0$$

$$\Rightarrow \hat{\beta}_2'' = \frac{\sum y_i x_{2i} - \hat{\alpha}'' \sum x_{2i}}{\sum x_{2i}^2}$$

با توجه به قضیه FWL می‌توان نتیجه گرفت که

$$\hat{\beta}_2' = \hat{\beta}_2''$$

C.

$x_{1i} = \gamma + \beta_{12} x_{2i} + x^*_{1i}$ (1)

We know $Cov(x^*_{1i}, x_{2i}) = 0$

$y_i = \alpha' + \beta_1' x^*_{1i} + \beta_2' x_{2i} + e_i$

$\rightarrow$ because $Cov(x^*_{1i}, x_{2i}) = 0$, β_2' is unaffected by the presence of x^*_{1i}, but that doesn't mean its unbiased.

(1) $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

(1) $\rightarrow y_i = \alpha + \beta_1 (\gamma + \beta_{12} x_{2i} + x^*_{1i}) + \beta_2 x_{2i} + u_i$

$\rightarrow y_i = (\alpha + \beta_1 \gamma) + \beta_1 x^*_{1i} + (\beta_1 \beta_{12} + \beta_2) x_{2i} + u_i$

We can conclude that $\boxed{\beta_2' \simeq \beta_1 \beta_{12} + \beta_2}$

$E(\beta_2') = E(\beta_1 \beta_{12} + \beta_2) = \underbrace{\beta_1 \beta_{12}}_{\text{Bias}} + \beta_2$
