

Homework 2 – Geometry of OLS
Due Feb 15th, 2007

QUESTION 1

- 1) Finish exercises 4, 5, 10, 11, and 13 from Chapter 3 of Greene (6th edition).
- 2) Finish exercises 7, 8, 9, and 10 from Chapter 4 of Greene (6th edition).

QUESTION 2

Question from Davidson and MacKinnon (2004), Chapter 2 (with some changes in notation):

Consider the following linear regression:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$$

where \mathbf{Y} is $n \times 1$, \mathbf{X}_1 is $n \times k_1$, and \mathbf{X}_2 is $n \times k_2$. Let $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ be the OLS parameter estimates from running this regression.

Now consider the following regressions, all to be estimated by OLS:

- a) $\mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- b) $\mathbf{P}_1\mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- c) $\mathbf{P}_1\mathbf{Y} = \mathbf{P}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- d) $\mathbf{P}_X\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- e) $\mathbf{P}_X\mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- f) $\mathbf{M}_1\mathbf{Y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- g) $\mathbf{M}_1\mathbf{Y} = \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- h) $\mathbf{M}_1\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- i) $\mathbf{M}_1\mathbf{Y} = \mathbf{M}_1\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$
- j) $\mathbf{P}_X\mathbf{Y} = \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}$

Here \mathbf{P}_1 projects orthogonally on to the span of \mathbf{X}_1 , and $\mathbf{M}_1 = \mathbf{I} - \mathbf{P}_1$. For which of the above regressions are the estimates of $\boldsymbol{\beta}_2$ the same as for the original regression? Why? For which are the residuals the same? Why?

QUESTION 3

EViews Exercise (This is very easy!)

This question uses artificial data to demonstrate the FWL theorem. Use the following commands to create an artificial dataset of size 100 in EViews:

```
gener x1 = 2 + nrnd
gener x2 = 4 + nrnd
gener e = nrnd
gener y = 2*x1 + 3*x2 + e
```

- a) Run the regression for y on x_1 and x_2 . You should get estimates that are reasonably close to the true values. Save the residual as `e_head`.
- b) If you need to run a regression with only x_1 as the independent variable, and you want the coefficient estimate on x_1 to be the same as that in a), what transformation do you need to perform on x_1 (without doing anything for y)? Show your regression

results.

c) Save the residual in b). Is it equal to e_{head} ? Why the mean of it is not zero?
Besides the transformation for x in b), what transformation do you need to perform on y to make the two residuals equal? Show your regression results.

QUESTION 4 – Freeman’s 1997 JLE Paper

- a) On page S152, why does the author claim that “while persons with higher earnings are more likely to volunteer than those with lower earnings capacity, the former generally volunteer fewer hours”?
- b) For the regression equation in Tables 3 and 4, discuss which OLS assumptions may be violated.
- c) There is a typo on page S158. Find it.
- d) For Table 6, compare and contrast the characteristics of individuals who volunteer when asked and those who volunteer on their own.
- e) For Table 6, is the sample used for the regression (3) the same as that for (4) or (5)?

QUESTION 5 – Hamermesh and Soss’s 1974 JPE Paper

- a) (*No credit*) Who is Schopenhauer? Why is he a less important German philosopher than Kant?
- b) On page 87, the authors claim that “within the United States, the use of suicide rates by states in a cross-section regression model could be dangerous because errors in the dependent variable are correlated with the true (observed) values of the income variables”. Explain his claim in more detail.
- c) On page 89, what does it mean intuitively when the discount rate r is infinity?
- d) On page 90, explain carefully what the authors mean by “this modification enables us to test the hypothesis that the apparent effect of increasing income over time on suicide rates merely reflects a trend in the latter.”
- e) The authors are not very clear about the structure of the data they use in section IV. Can you explain it clearer? (Do we have observations on individuals? Or groups? Do we observe the same individuals over time? ...)
- f) On page 93, explain what the authors mean by “if there is a negative correlation between the percentage Catholic and occupational status within each narrow jurisdiction, the results presented in table 4 will be biased against observing the expected negative relation between suicide and occupational status.”
- g) On page 94, explain the claim that “given the lower incidence of suicide among Catholics, a failure to control for their proportion in a state’s population would bias the coefficient of income negatively.”
- h) On page 96, why do the authors claim that “current incomes are a good proxy for lifetime incomes here and may well perform better”?

QUESTION 6 – Arcelus and Meltzer’s 1975 APSR Paper

- a) On page 1234, when defining the habitual voters do the authors assume the effect of the two realignments to be permanent or temporary?
- b) In all the equations, who does the “ Δ ” sign mean intuitively (it has no mathematical meaning though)?

- c) For equation (2), is it a problem for some voters to belong to more than one category (e.g. a voter who is both partisan and affected by aggregate economic performance)?
- d) Explain carefully the purpose of having the last three interaction terms in equation (3). What do the coefficients on them tell you?
- e) On page 1236, the authors claim that “currently, in non-presidential years, 42.6 per cent of the eligible voters are partisans who vote in congressional elections.” Where does the number 42.6 come from?

QUESTION 7 – Meltzer and Vellrath’s 1975 JLE Paper

- a) On page 787 the Empirical Results section, the authors claim that “there is a group of consistent voters, the habitual voters, consisting of approximately 50 per cent of the eligible voters.” Where does the number 50 come from?
- b) On page 789, based on which results do the authors claim that “the transfers appear to shift votes from the Democrats to the Republicans, on average, and taxes shift votes from the Republicans mainly to non-voters”?
- c) Explain how the numbers in Table 2 are calculated.
- d) Explain how the numbers in Table 4 are calculated.
- e) Explain intuitively why the authors include the lagged dependent variable in Table 5.