

Econometrics A

Problem Set #5

Date Due: Thursday, May 22

1. Stock and Watson, Exercises 7.9, E7.3, E8.1
2. Consider the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U .$$

Let $X = (1, X_1, X_2, X_3)'$. Suppose

$$E[XU] = 0$$

and that X is not perfectly colinear. Suppose further that $E[Y^4] < \infty$ and $E[X_j^4] < \infty$ for $1 \leq j \leq 3$. Using a large sample of i.i.d. observations from (Y, X_1, X_2, X_3) , you estimate this equation using OLS. Let $\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_3)$ denote the resulting estimate of $\beta = (\beta_0, \dots, \beta_3)$.

- (a) Let $\theta = 2\beta_1 - \beta_3$. Is $\hat{\theta}_n = 2\hat{\beta}_1 - \hat{\beta}_3$ a consistent estimate of θ ? Is it an unbiased estimate of θ ?
 - (b) Express $\text{Var}[\hat{\theta}_n]$ in terms of $\text{Var}[\hat{\beta}_1]$, $\text{Var}[\hat{\beta}_3]$, and $\text{Cov}[\hat{\beta}_1, \hat{\beta}_3]$.
 - (c) How would you test the null hypothesis $H : \theta = 1$ versus the alternative $K : \theta \neq 1$ at the 5% significance level? In particular, write down your test statistic, your critical value, and rule you would use to determine whether or not to reject the null hypothesis.
 - (d) What is the p -value for your test?
3. Consider the following regression:

$$Y = X'\beta + U ,$$

where $E[U|X] = 0$ (remember that this is a stronger assumption than our usual assumption that $E[XU] = 0$). Suppose Y is binary.

- (a) What is $P\{Y = 1|X\}$?
- (b) What is $\text{Var}[U|X]$? Is it reasonable to assume that U is homoskedastic?
- (c) Is it possible for the fitted value of Y to lie outside $[0, 1]$?

4. Download the dataset for this problem from the TA's webpage. The dataset has 4 variables in the following order: *college* (an indicator for completing college), *hs* (an indicator for completing high school, but not college), *wage* (average hourly wage) and *fem* (an indicator for female). In order to load the data into Stata, you may need to increase the memory. To do this, type

```
set mem 10m
```

in Stata. This will increase the memory from the default of 1 mb to 10 mb.

- (a) Consider the following model of the determinants of income:

$$wage = \beta_0 + \beta_1 fem + \beta_2 college + \beta_3 hs + U .$$

Suppose that the regressors are not perfectly colinear.

- i. What is the interpretation of U ? Is it uncorrelated with the regressors?
- ii. Interpret each of the coefficients in the model.
- iii. For the rest of part (a) and part (b) below, suppose U is uncorrelated with the regressors and that the fourth moments of the dependent variable and each of the regressors exist. Estimate the model using OLS and construct a 95% confidence interval for the effect of being female on income. What coverage property does the interval satisfy?
- iv. Let $\theta = \beta_2 - \beta_3$ and $\hat{\theta}_n = \hat{\beta}_2 - \hat{\beta}_3$.
 - A. Interpret θ .
 - B. Under what conditions will $\hat{\beta}_2$ and $\hat{\beta}_3$ be unbiased estimators for β_2 and β_3 , respectively? Under these conditions, will $\hat{\theta}_n$ be an unbiased estimator of θ ?
 - C. Are $\hat{\beta}_2$ and $\hat{\beta}_3$ consistent estimators of β_2 and β_3 , respectively? Is $\hat{\theta}_n$ a consistent estimator of θ ?
 - D. Derive an expression for the variance of $\hat{\theta}_n$ in terms of $\text{Var}[\hat{\beta}_2]$, $\text{Var}[\hat{\beta}_3]$ and $\text{Cov}[\hat{\beta}_2, \hat{\beta}_3]$. Using your results, what is your estimated standard deviation (i.e., standard error) for $\hat{\theta}_n$?
- v. Without running a new regression, test $H_0 : \beta_2 = \beta_3$ versus $H_1 : \beta_2 > \beta_3$ at the 5% significance level. Be careful to explain how you construct the test. Explain the null and alternative hypotheses in words.

- (b) Use the dataset to estimate the following model,

$$wage = \gamma_0 + \gamma_1 fem + \gamma_2 college + \gamma_3 (hs + college) + U.$$

To do this regression, you will have to create a new variable that is the sum of *hs* and *college*.

- i. How do the coefficients $\gamma_0, \dots, \gamma_3$ relate to the coefficients β_0, \dots, β_3 defined in part (a)?
 - ii. Test $H_0 : \gamma_2 = 0$ versus $H_1 : \gamma_2 > 0$ at the 5% significance level. How do the results of your test compare to what you found in part (v) of part (a) above? Explain briefly.
- (c) Consider the following model of the determinants of income:

$$\begin{aligned} \text{wage} = \\ \beta_0 + \beta_1 \text{fem} + \beta_2 \text{college} + \beta_3 \text{fem} \times \text{college} + \beta_4 \text{hs} + \beta_5 \text{fem} \times \text{hs} + U . \end{aligned}$$

Suppose that the regressors are not perfectly colinear.

- i. Interpret each of the coefficients in the model.
- ii. For the rest of part (c), suppose that U is uncorrelated with each of the regressors and that the fourth moments of the regressors and dependent variable exist. Estimate the model using OLS. According to your results, what is the estimated effect on income of $\text{college} = 1$ versus $\text{hs} = 1$ for men? Construct a 95% confidence interval for this effect. What coverage property does the interval satisfy?
- iii. Consider the null hypothesis that the effect on income of $\text{college} = 1$ versus $\text{hs} = 1$ is the same for men and women, versus the alternative that they are different for men and women. Formally state the null and alternative hypotheses. Conduct the test at the 10% level.
- iv. Consider the null hypothesis that there are no interactions between gender and education. Formally state the null and alternative hypotheses. Conduct the test at the 10% level.