Introductory Econometrics I – Midterm Exam

Yingjie Feng

School of Economics and Management

Tsinghua University

April 11, 2022

Contents

Notes:

- Duration of examination: *120 minutes*.
- Please write your name and student ID clearly on the first page of the answer book.
- Please do not open the exam paper until the proctors ask you to do so.
- Please answer *all* questions. Feel free to use either English or Chinese.
- Please start each question on a separate page.
- Answers without proper justification will *not* receive (partial) credit.
- Please return the exam question book and answer book at the end of the exam.

1 Question 1 (25 points)

Consider the following regression model:

$$y = \beta_0 + \beta_1 x + u, \qquad \beta_0 = 0.$$

That is, in this special case the intercept β_0 is zero. Suppose that (the other four) Gauss-Markov assumptions also hold: (i) A random sample $\{(x_i, y_i) : 1 \leq i \leq n\}$ is available; (ii) $\{x_i : 1 \leq i \leq n\}$ are not all the same; (iii) $\mathbb{E}[u|x] = 0$; and (iv) $\mathbb{V}[u|x] = \sigma^2$. The following estimator of β_1 has been proposed.

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}.$$

Assume $\sum_{i=1}^n x_i \neq 0$ and $\mathbb{E}[x] \neq 0$ so that the above estimator is always well defined.

1. [5 points] Show $\tilde{\beta}_1$ is unbiased in this *special* case. [Hint: As in class, take expectation conditional on $\{x_i : 1 \leq i \leq n\}$.]

   - **Solution**: By Assumption SLR.4, conditional on $\{x_i : 1 \leq i \leq n\}$,

   $$\mathbb{E}[\tilde{\beta}_1|x] = \beta_1 + \frac{\sum_{i=1}^n \mathbb{E}[u_i|x]}{\sum_{i=1}^n x_i} = \beta_1 + 0$$

2. [5 points] Derive the variance of $\tilde{\beta}_1$ (conditional on $\{x_i : 1 \leq i \leq n\}$).

   - **Solution**:

   $$\mathbb{V}[\tilde{\beta}_1|x] = \mathbb{V}\Big[\frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n x_i}\Big] = \frac{\sum_{i=1}^n \mathbb{V}[u_i|x]}{(\sum_{i=1}^n x_i)^2} = \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2}$$

3. [5 points] Run a regression of y_i on 1 and x_i as usual (that is, including an intercept in the regression). Denote by $\hat{\beta}_1$ the OLS estimator of the slope β_1 on x. Is $\hat{\beta}_1$ unbiased? What is the variance of $\hat{\beta}_1$ (conditional on $\{x_i : 1 \leq i \leq n\}$)? You *do not* need to give a proof.

   - **Solution**: From the class, $\hat{\beta}_1$ is unbiased and the variance of $\hat{\beta}_1$ is

   $$\mathbb{V}[\hat{\beta}_1|x] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

4. [5 points] Do you think we can apply the Gauss-Markov theorem to conclude the variance of

$\hat{\beta}_1$ (conditional on $\{x_i : 1 \le i \le n\}$) is no greater than that of $\tilde{\beta}_1$ *in general*? Explain why or why not. [Hint: Do you think $\tilde{\beta}_1$ is still an unbiased estimator when $\beta_0 \neq 0$?]

- **Solution**: No. When β_0 is nonzero, we have

$$\mathbb{E}[\tilde{\beta}_1|x] = \frac{n\beta_0}{\sum_{i=1}^n x_i} + \beta_1 + \frac{\sum_{i=1}^n \mathbb{E}[u_i|x]}{\sum_{i=1}^n x_i} = \beta_1 + \frac{n\beta_0}{\sum_{i=1}^n x_i},$$

which is not β_1 unless $\beta_0 = 0$. So $\tilde{\beta}_1$ is generally not an unbiased estimator of β_1. Then, we cannot apply Gauss-Markov theorem since the theorem only says the OLS estimators have the smallest variance within the family of linear unbiased estimators. In fact, comparing the two variances derived previously, it is possible that $\mathbb{V}[\tilde{\beta}_1|x] < \mathbb{V}[\hat{\beta}_1|x]$.

5. [5 points] Suppose (only in this part) that the true intercept $\beta_0 \neq 0$ (so our initial setup is incorrect). What is the probability limit of $\tilde{\beta}_1$ as the sample size $n \to \infty$ in this case? Is $\tilde{\beta}_1$ consistent?

- **Solution**: It is inconsistent since by law of large numbers

$$\tilde{\beta}_1 - \beta_1 = \frac{\beta_0}{\frac{1}{n}\sum_{i=1}^n x_i} + \frac{n^{-1}\sum_{i=1}^n u_i}{n^{-1}\sum_{i=1}^n x_i} \to_p \beta_0/\mathbb{E}[x] + 0$$

2 Question 2 (25 points)

Consider an equation to explain salaries (in Chinese *yuan*) of CEOs in terms of annual firm *sales* (in Chinese *yuan*), return on equity (*roe*, in percentage form), and return on the firm's stock (*ros*, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u$$

1. [5 points] The OLS estimate of β_1 is 0.28. How do you interpret this estimate?

- **Solution**: Holding other factors fixed, 1% increase in *sales* will lead to 0.28% increase in CEO salary (elasticity of *salary* with respect to *sales*).

2. [5 points] If we measure *sales* in 1,000 *yuan* rather than *yuan* (that is, define a new variable $sale1k = sale/1000$), does the slope estimate $\hat{\beta}_1$ change? Does the intercept estimate $\hat{\beta}_0$ change (in general)? Explain why.

   - **Solution**: $\hat{\beta}_1$ does not change, but the intercept $\hat{\beta}_0$ will change, since $\log(sales/1000) = \log(sales) - \log(1000)$ and this transformation only affects the intercept of the regression (β_1 as the elasticity is free of units of measurement).

3. [5 points] State the null hypothesis that, after controlling for sales, the effect of *roe* on CEO salary is the same as that of *ros*. State the alternative hypothesis that each percentage point increase in *roe* has a larger (more positive) effect on CEO salary than that in *ros*.

   - **Solution**: $H_0 : \beta_2 = \beta_3$, $H_1 : \beta_2 > \beta_3$.

4. [5 points] Let $\theta = \beta_2 - \beta_3$. Write the regression equation involving β_0, β_1 and β_3 and θ to directly obtain the estimate and standard error of θ.

   - **Solution**:

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + (\beta_2 - \beta_3)roe + \beta_3(ros + roe) + u$$

5. [5 points] Based on your answer to part 4, describe in detail how to test the null hypothesis against the alternative hypothesis you state in part 3 at the 1% level using a t test.

   - **Solution**: t test: reject H_0 if the t-ratio= $\hat{\theta}/se(\hat{\theta})$ is greater than 99% percentile of t distribution with $n - 4$ degrees of freedom.

3 Question 3 (50 points)

We have a dataset on workers' wages and other features:

- $lwage = log$(monthly wage)
- $educ$ = years of education

- $pareduc$ = total amount of both parents' education
- $exper$ = years of working experience
- $tenure$ = years with current employer

1. Consider the following regression model

$$lwage = \beta_0 + \beta_1 educ + \beta_2 pareduc + u.$$

One gets the following estimate

$$\widehat{lwage} = 5.935 + 0.047 educ + 0.011 pareduc + u.$$

(a) [5 points] We also know that the sample correlation coefficient between $pareduc$ and $educ$ is 0.452. Do you think the likely bias in $\tilde{\beta}_1$ obtained from a simple regression of $lwage$ on $educ$ is positive or negative? Explain why.

- **Solution**: Based on the relation $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$, $\mathbb{E}[\tilde{\beta}_1] > \beta_1$.

(b) [5 points] In the dataset we actually have both mother's education ($meduc$) and father's education ($feduc$). Do you think we can add both $meduc$ and $feduc$ to the above regression and simultaneously estimate the effects of $meduc$, $feduc$ and $pareduc$? Explain why or why not.

- **Solution**: No. We run into the perfect collinearity issue: $pareduc = feduc + meduc$.

2. Now, consider the regression model

$$lwage = \beta_0 + \beta_1 educ + \beta_2 educ \times pareduc + \beta_3 exper + \beta_4 tenure + u.$$

The following is the regression table given by Stata, where $educ\_pareduc = educ \times pareduc$. Answer the following questions.

Note: Throughout this part, you can give answers that may involve sums, products, quotients or square root of known values; e.g., you *do not* have to actually calculate a value.

Source	SS	df	MS			
				Number of obs	=	722
				F(4, 717)	=	$\boxed{36.44}$
Model	21.4253649	4	5.35634121	Prob > F	=	0.0000
Residual	105.386551	717	.146982637	R-squared	=	(a)
				Adj R-squared	=	0.1643
Total	126.811916	721	.175883378	Root MSE	=	(b)

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0467522	.0104767	4.46	0.000	.0261835	(c)
educ_pareduc	.000775	.0002107	3.68	0.000	.0003612	.0011887
exper	.018871	.0039429	4.79	0.000	.0111299	.026612
tenure	.0102166	.0029938	(d)	0.001	.0043391	.0160942
_cons	(e)	.1295593	43.58	0.000	5.392158	5.90088

(a) [15 points] There are five missing values in the table (labeled (a), (b), (c), (d) and (e)). Fill them in. [Hint: the 97.5% percentile of t distribution with 717 degrees of freedom is 1.96.]

- **Solution**: (a) $SSE/SST = 21.425/126.812$ or $1-(1-\bar{R}^2)*(n-k-1)/(n-1) = 1-(1-0.1643)*717/721$; (b) $\sqrt{\frac{1}{n-k-1}SSR} = \sqrt{0.1470}$; (c) $\hat{\beta}_1 + 1.96se(\hat{\beta}_1) = 0.047 + 1.96 * 0.010$; (d) $t(\hat{\beta}_4) = \hat{\beta}_4/se(\hat{\beta}_4) = 0.01/0.00299$; (e) $\hat{\beta}_0 = t(\hat{\beta}_0) * se(\hat{\beta}_0) = 43.58 * 0.1296$

(b) [10 points] There is an F statistic at the top-right corner (in box). What are the null and alternative hypotheses of this test? Explain how this F statistic is calculated.

- **Solution**: $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, $H_1 : H_0$ is not true.

$$F = \frac{R^2/4}{(1-R^2)/717}$$

or

$$F = \frac{SSE/4}{SSR/717} = \frac{21.425/4}{105.387/717}$$

(c) [10 points] How do you interpret the coefficient on *educ*? What is the return to education if one's parents both have high school education, that is, $pareduc = 24$?

- **Solution**: If one's parents did not receive any education, holding other factors fixed, the return to each year of education is 4.7%.

$0.047 + 0.00078 * 24 = 0.066$. That is, if one's education increases by 1 year, wage increases by 6.6%.

3. [5 points] Finally, add *pareduc* to the regression model in part 2 as a separate variable. We obtain the following result:

$$\widehat{lwage} = \underset{(0.38)}{4.94} + \underset{(0.027)}{0.097}\, educ + \underset{(0.017)}{0.033}\, pareduc - \underset{(0.0012)}{0.0016}\, educ \cdot pareduc + \underset{(0.004)}{0.020}\, exper + \underset{(0.003)}{0.010}\, tenure$$

Standard errors are in parentheses. Does the estimated return to education depend positively on parent education now? Test the null hypothesis that the return to education does not depend on parent education at the 5% level. [Hint: the critical value is 1.96.]

- **Solution**: No, the sign of the interaction term is negative, indicating the estimated return to education depends negatively on parents' education. However, the t test statistic is $0.0016/0.0012 = 4/3 \approx 1.3 < 1.96$. Thus, we fail to reject the null hypothesis.