Economics 471: Introductory Econometrics

Department of Economics, Finance and Legal Studies

University of Alabama

Fall 2021

Midterm II Key

The exam consists of three questions on four pages. Each question is of equal value.

1. Suppose true data generating process is $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + v_i$, where $E(v_i|x_i) = 0$, but you instead run the model $y_i = \beta_0 + \beta_1 x_i + u_i$.

   (a) What Gauss-Markov assumption(s) are violated in your model?

   (b) What is the estimator of β_1 for your model (i.e., $\widehat{\beta}_1$)?

   (c) What is $E\left(\widehat{\beta}_1 | X = x_i\right)$ from your estimator in part (b)?

   (d) Give the equation for the bias of your estimator in part (b).

   (e) Under what conditions will your estimator from part (b) be unbiased?

$u_i = \beta_2 x_i^2 + v_i$

(a) $E(u_i|x_i) = E(\beta_2 x_i^2 + v_i | x_i) = \beta_2 x_i^2 \neq 0$ (3)

$E(u_i) = E(\beta_2 x_i^2 + v_i) = \beta_2 x_i^2$ (2)

$y_i = \beta_0 + \beta_1 x_i + u_i \neq E(y_i|x_i)$ (1)

(b) $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(c) $$E(\hat{\beta}_1 | X) = E\left(\frac{\sum_{i=1}^{n} (\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + v_i)(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \Big| X \right)$$

$$= \beta_1 + \beta_2 \frac{\sum_{i=1}^{n} x_i^2 (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(d) $\text{bias}(\hat{\beta}_1) = E(\hat{\beta}_1 | x) - \beta_1$

$$= \beta_2 \frac{\sum_{i=1}^{n} x_i^2 (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(e) $\beta_2 = 0$

$\sum_{i=1}^{n} x_i^2 (x_i - \bar{x}) = 0$

2. Suppose our model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + u$ and we wish to test the null that y is solely a linear function of x_1 and x_2. For this test, we plan to use the $F-$statistic discussed in class

$$F = \frac{(R_U^2 - R_R^2)/q}{(1 - R_U^2)/(n-k-1)}$$

   (a) Write down the null hypothesis in terms of the coefficients.

   (b) For hypothesis in part (a), define each component on the right hand side of F.

   (c) What is the distribution of this test statistic (be sure to list the degrees of freedom)? Draw this distribution.

   (d) What is the range of the test statistic? Why?

   (e) Using this $F-$statistic, derive the $F-$statistic in terms of the SSR formulation. Show your work.

(a) $H_0: \beta_{11} = \beta_{22} = \beta_{12} = 0$

R: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

U: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + u$

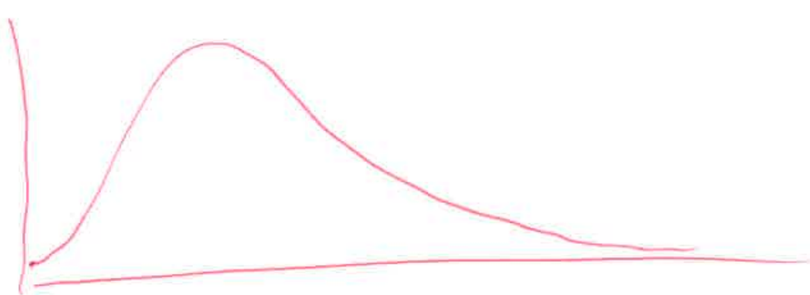
(b) R_u^2 - R^2 from reg U

R_R^2 - R^2 from reg R

$q = 3$ (# of restrictions

$(n-k-1) = n-6$

(c) $F \sim F_{3,(n-6)}$



(d) $0 \leq R_R^2 \leq R_u^2 \leq 1 \Rightarrow \underline{F \geq 0}$

(e) $R_R^2 = 1 - \frac{SSR_R}{SST}$

$R_u^2 = 1 - \frac{SSR_u}{SST}$

$$F = \frac{\left[\left(1 - \frac{SSR_u}{SST}\right) - \left(1 - \frac{SSR_R}{SST}\right)\right] / q}{\left[1 - \left(1 - \frac{SSR_u}{SST}\right)\right] / (n-k-1)}$$

$$= \frac{\left(\frac{SSR_R}{SST} - \frac{SSR_u}{SST}\right) / q}{\left(\frac{SSR_u}{SST}\right) / (n-k-1)}$$

$$= \frac{(SSR_R - SSR_u) / q}{SSR_u / (n-k-1)}$$

$= F$

3. Consider the gretl output below relating test scores (testscores) to hours of homework per day (homework), class size (classsize), hours of class (hrsofclass) per week, and previous test scores (prevtestscores). With the output from these two models, answer the questions on the following page:

Model 1: OLS, using observations 1–3733
Dependent variable: testscores

	Coefficient	Std. Error	t-ratio	p-value
const	49.8379	0.298027	167.2	0.0000
homework	4.01135	0.393603	10.19	0.0000

Mean dependent var	52.43538	S.D. dependent var	9.566599
Sum squared resid	332301.3	S.E. of regression	9.437423
R^2	0.027084	Adjusted R^2	0.026823
$F(1, 3731)$	103.8635	P-value(F)	4.47e–24
Log-likelihood	−13675.30	Akaike criterion	27354.60
Schwarz criterion	27367.05	Hannan–Quinn	27359.03

Model 2: OLS, using observations 1–3733
Dependent variable: testscores

	Coefficient	Std. Error	t-ratio	p-value
const	7.95590	0.569671	13.97	0.0000
homework	0.875984	0.200163	4.376	0.0000
classsize	0.0164890	0.0110339	1.494	0.1352
hrsofclass	−0.0684837	0.0784510	−0.8729	0.3827
prevtestscores	0.831823	0.00794525	104.7	0.0000

Mean dependent var	52.43538	S.D. dependent var	9.566599
Sum squared resid	83841.82	S.E. of regression	4.742337
R^2	0.754527	Adjusted R^2	0.754263
$F(4, 3728)$	2864.750	P-value(F)	0.000000
Log-likelihood	−11104.92	Akaike criterion	22219.84
Schwarz criterion	22250.97	Hannan–Quinn	22230.92

(a) Interpret the coefficient on homework from model 2. Why does this differ from that in model 1?

(b) Test the null hypothesis that the intercept is zero (i.e., $H_0 : \alpha = 0$) in model 2.

(c) Test the null hypothesis that the additional three variables (classsize, hrsofclass and prevtestscores) are jointly irrelevant in model 2.

(d) Test the validity of the regression in model 1.

(e) Suppose we divided test scores by 100 (for all observations). What will happen to SST? What will happen to R^2?

(a) a one hour increase in HW/day ⇒ 0.875 additional points (expected) on the exam

model 1 is biased upwards due to omitted variable bias

(b) $H_0: \alpha = 0$

$H_1: \alpha \neq 0$

$$t = \frac{\hat{\alpha} - 0}{se(\hat{\alpha})} = \frac{7.9559 - 0}{0.569671} = 13.9772$$

⇒ reject H_0 (p-value = 0.0000)

(c) can use R^2 or SSR version $H_0: \beta_2 = \beta_3 = \beta_4 = 0$

$$F = \frac{(R^2_u - R^2_R)/q}{(1-R^2_u)/(n-k-1)} = \frac{(0.7345 - 0.027)/3}{(1-0.7345)/(3733-5)}$$

which is a big number clearly, so we expect to reject H_0

(d) $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

$$F = \frac{(R^2_u - R^2_R)/q}{(1-R^2_u)/(n-k-1)} = \frac{(0.027084 - 0)/1}{(1-0.027084)/(3733-2)}$$

$= 103.8635$

p-value $= 4.47e{-24} \Rightarrow$ reject H_0

(can also use SSR version, but the value for F & p-value are known here)

(e) $y_i^* = \dfrac{y_i}{100}$

$$\overline{y_i^*} = \frac{1}{n}\sum_{i=1}^{n} y_i^* = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{100} = \frac{1}{100}\cdot\frac{1}{n}\sum_{i=1}^{n} y_i$$

$$= \frac{\bar{y}}{100}$$

$$SST^* = \sum_{i=1}^{n}\left(y_i^* - \overline{y_i^*}\right)^2 = \underline{\frac{1}{100^2} SST}$$

$$SSE^* = \sum_{i=1}^{n}\left(\hat{y}_i^* - \overline{y_i^*}\right)^2 = \frac{1}{100^2} SSE$$

$$R^2 = \frac{SSE^*}{SST^*} = \frac{\frac{1}{100^2} SSE}{\frac{1}{100^2} SST} = \frac{SSE}{SST} = R^2$$

<u>no change in R^2</u>