Economics 471: Introductory Econometrics

Department of Economics, Finance and Legal Studies

University of Alabama

Fall 2021

Final Exam Key

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

1. Suppose true data generating process is $y_i = \beta x_i + u_i$, where $E(u_i|x_i) = 0$ and $V(u_i|x_i) = \sigma_i^2$.

   (a) Derive the least-squares estimator of β.

   (b) Show that the estimator from part (a) is unbiased.

   (c) Derive the variance of the estimator from part (a) for the case where $V(u_i|x_i) = \sigma_i^2 = \sigma^2$, where σ^2 is known.

   (d) Derive the variance of the estimator from part (a) for the case where $V(u_i|x_i) = \sigma_i^2$, where σ_i^2 is known.

   (e) Derive the variance of the estimator from part (a) for the case where $V(u_i|x_i) = \sigma_i^2$, where σ_i^2 is unknown (i.e., the "robust", White [1980] version).

(a) $$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta} x_i)^2$$

$$\frac{\partial}{\partial \beta} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta} x_i) x_i = 0$$

$$\sum_{i=1}^{n} y_i x_i = \hat{\beta} \sum_{i=1}^{n} x_i^2$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2}$$

(b) $$\hat{\beta} = \frac{\sum_{i=1}^{n} (\beta x_i + u_i) x_i}{\sum_{i=1}^{n} x_i^2} = \beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2}$$

$$E(\hat{\beta}|x_i) = E\left(\beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \middle| x_i\right)$$

$$= \beta + \frac{\sum_{i=1}^{n} x_i E(u_i|x_i)}{\sum_{i=1}^{n} x_i^2} \leftarrow 0$$

$$= \beta$$

(c) $$V(\hat{\beta}|x_i) = V\left(\beta + \frac{\sum_{i=1}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \middle| x_i\right)$$

$$\overset{iid}{=} \frac{\sum_{i=1}^{n} x_i^2 V(u_i|x_i)}{\left(\sum_{i=1}^{n} x_i^2\right)^2}$$

$$\overset{\sigma_i^2=\sigma^2}{=} \frac{\sum_{i=1}^{n} x_i^2 \sigma^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

(d) $$V(\hat{\beta}|x_i) = \sum_{i=1}^{n} x_i^2 \sigma_i^2 \Big/ \left(\sum_{i=1}^{n} x_i^2\right)^2$$

(e) $$V(\hat{\beta}|x_i) = \sum_{i=1}^{n} x_i^2 \sigma_i^2 \Big/ \left(\sum_{i=1}^{n} x_i^2\right)^2$$

$$\hat{V}(\hat{\beta}|x_i) = \sum_{i=1}^{n} x_i^2 \hat{u}_i^2 \Big/ \left(\sum_{i=1}^{n} x_i^2\right)^2$$

2. Suppose our model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, where $E(u|x_1, x_2) = 0$. With this information, answer the following (be specific):

(a) Derive the estimator for β_0.

(b) Using the formula from class, write down the estimator for β_1.

(c) Using the formula from class, write down the variance of the estimator from part (b).

(d) Suppose $corr(x_1, x_2) = 1$, what will happen to the variance of the estimator from part (c)?

(e) For the null $H_0 : \beta_1 = \beta_2 = 0$, give the test statistic and the distribution of the test statistic under the null hypothesis. Draw this distribution and indicate the rejection region.

(a) $$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$$

$$\frac{\partial}{\partial \beta_0} = -\frac{2}{n} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}) = \frac{0}{n}$$

$$\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_{1i} - \hat{\beta}_2 \frac{1}{n}\sum_{i=1}^{n} x_{2i} = 0$$

$$\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 = 0$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

(b) $$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i \hat{r}_{1i}}{\sum_{i=1}^{n} \hat{r}_{1i}^2}$$

$$x_{1i} = \delta_0 + \delta_2 x_{2i} + r_{1i}$$

$$\hat{r}_{1i} = x_{1i} - \hat{\delta}_0 - \hat{\delta}_2 x_{2i}$$

(c) $$V(\hat{\beta}_1 | x) = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x}_1)^2 (1 - R_1^2)}$$

where R_1^2 is the R^2 from

$$x_{1i} = \gamma_0 + \gamma_2 x_{2i} + r_{1i}$$

(d) if $\text{CORR}(x_1, x_2) = 1 \Rightarrow R_1^2 = 1$

$$\Rightarrow V(\hat{\beta}_1 | x) = \infty$$

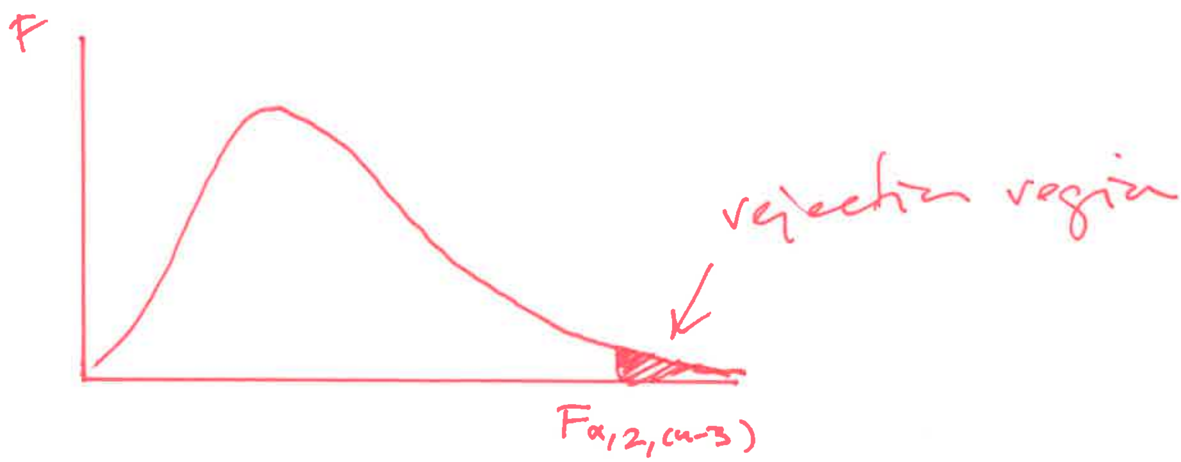
(e) $H_0: \beta_1 = \beta_2 = 0$

H_1: H_0 is not true

$$F = \frac{(SSR_R - SSR_U)/2}{SSR_U/(n-3)} \sim F_{2,(n-3)}$$

or

$$F = \frac{(R^2_U - R^2_R)/2}{(1 - R^2_U)/(n-3)} \sim F_{2,(n-3)}$$



3. We are interested in studying the starting wage (Y) for union ($D = 1$) and non-union ($D = 0$) members. Suppose our model is $Y = \alpha + \delta D + U$, where it is assumed that $E(U|D) = 0$. Using conditional expectations, answer the following:

(a) Define the conditional expectation for each group.

(b) Define the base group.

(c) What is the marginal impact of going from group 0 to group 1?

(d) Suppose we had a random sample of data, what would be a good estimator for α?

(e) Suppose we had a random sample of data, what would be a good estimator for δ?

(a) $E(Y|D=1) = E(\alpha+\delta D+U|D=1) = \alpha+\delta$

$E(Y|D=0) = E(\alpha+\delta D+V|D=0) = \alpha$

(b) $D=0$ non-union members

(c) $E(Y|D=1) - E(Y|D=0) = (\alpha+\delta)-\alpha = \delta$

(d) $\hat{\alpha} = \hat{E}(Y|D=0) = \overline{y_0}$

$\overline{y_0}$ is the sample mean of y for those in group 0

(e) $\hat{\delta} = \hat{E}(Y|D=1) - \hat{E}(Y|D=0)$

$= \overline{y_1} - \overline{y_0}$

$\overline{y_1}$ is the sample mean of y for those in group 1

4. 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, previous test scores (prevtestscores), sex of the student (sex), sex of the teacher (teachersex) and mother's education (mothereduc). Note that sex and teachersex are binary (dummy) variables (equal to 1 if female and zero otherwise). With this information, answer the following:

Model 1: OLS, using observations 1–3733
Dependent variable: testscores
Heteroskedasticity-robust standard errors, variant HC1

	Coefficient	Std. Error	t-ratio	p-value
const	7.75402	0.58451	13.270	0.0000
homework	0.86239	0.20863	4.133	0.0000
hrsofclass	−0.06246	0.08029	−0.778	0.4366
classsize	0.01511	0.01175	1.285	0.1987
prevtestscores	0.81741	0.00796	102.600	0.0000
sex	0.05294	0.15554	0.340	0.7336
teachersex	0.08703	0.15557	0.559	0.5759
mothereduc	0.30608	0.05172	5.918	0.0000

Mean dependent var	52.43538	S.D. dependent var	9.566599
Sum squared resid	83031.55	S.E. of regression	4.721266
R^2	0.756899	Adjusted R^2	0.756442
$F(7, 3725)$	2025.534	P-value(F)	0.000000
Log-likelihood	−11086.80	Akaike criterion	22189.59
Schwarz criterion	22239.39	Hannan–Quinn	22207.31

(a) Interpret the coefficient on hours of class.

(b) Interpret the coefficient on teachersex.

(c) Suppose we wish to test the null that homework is insignificant (i.e., $H_0 : \beta_{homework} = 0$). Give the test statistic and the distribution of the test statistic under the null hypothesis. Draw this distribution and indicate the rejection region.

(d) Suppose we wish to test the null that sex is insignificant (i.e., $H_0 : \beta_{sex} = 0$). Give the test statistic and the distribution of the test statistic under the null hypothesis. Draw this distribution and indicate the rejection region.

(e) Suppose we wanted to see if female students performed better with female teachers. What would you need to add to this model to examine this result?

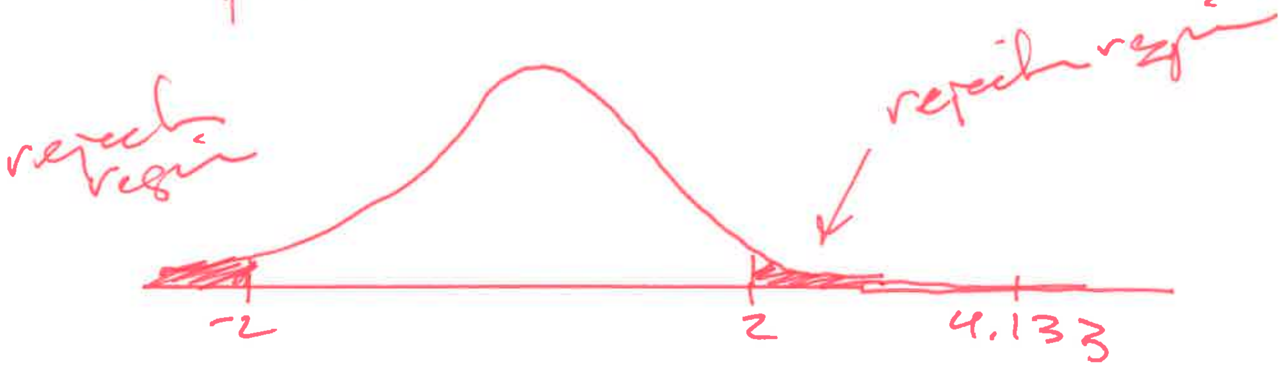
(a) an increase of 1 hour per week of class $\Rightarrow$ predicted <u>decrease</u> in test scores by 0.06246 points

(b) going from a male to a female teacher, all else constant, will increase the predicted test score by 0.08703 points

(c) $H_0: \beta_{hw} = 0$

$H_1: \beta_{hw} \neq 0$

$$t = \frac{\hat{\beta}_{hw} - 0}{se(\hat{\beta})} = \frac{0.86239 - 0}{0.20863} = 4.133$$

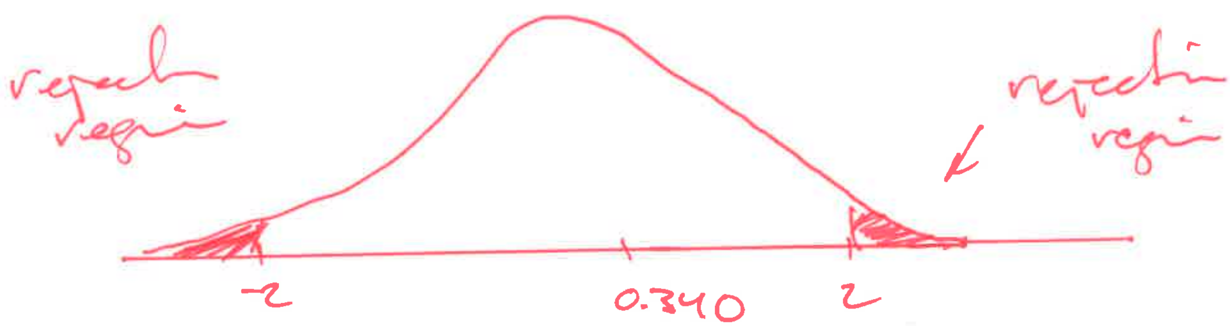


note: w/ robust se, the null dist may not be close to $t_{(n-8)}$

(d) $H_0: \beta_{sex} = 0$

$H_1: \beta_{sex} \neq 0$

$$t = \frac{\hat{\beta}_{sex} - 0}{se(\hat{\beta}_{sex})} = \frac{0.05294 - 0}{0.15554} = 0.340$$



note: w/ robust se, the null dist may not be close to $t_{(n-8)}$

(e) we would want to interact the sex & teach sex dummy variables

$\beta_{st}\ sex \cdot teachsex$