Economics 471: Introductory Econometrics

Department of Economics, Finance and Legal Studies

University of Alabama

Spring 2022

Midterm II — Key

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

1. Consider a random sample of data $\{x_{1i}, x_{2i}, y_i\}_{i=1}^{n}$ and the model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$, where $E(u_i|x_{1i}, x_{2i}) = 0$. We know that an estimator of β_1 is

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} \widehat{r}_{1i} y_i}{\sum_{i=1}^{n} \widehat{r}_{1i}^2}$$

and the conditional variance of that estimator is

$$\widehat{V}\left(\widehat{\beta}_1 | x_{1i}, x_{2i}\right) = \frac{\widehat{\sigma}^2}{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2 (1 - R_1^2)}.$$

With this information, answer the following questions:

(a) What model is used to estimate r_{1i}?

(b) For the model in part (a), derive the estimator of the intercept parameter.

(c) For the model in part (a), derive the estimator of the slope parameter.

(d) Write down the estimator for the error variance term $\widehat{\sigma}^2$.

(e) Suppose x_1 and x_2 are uncorrelated, what does the conditional variance simplify to (be specific)?

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

(b) $\sum_{i=1}^{n} \hat{r}_{1i}^2 = \sum_{i=1}^{n} (x_{1i} - \hat{\gamma}_0 - \hat{\gamma}_2 x_{2i})^2$

$\Rightarrow \hat{\gamma}_0 = \bar{x}_1 - \hat{\gamma}_2 \bar{x}_2$

(c) $\Rightarrow \hat{\gamma}_2 = \dfrac{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)(x_{1i} - \bar{x}_1)}{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2}$

(d) $\hat{\sigma}^2 = \dfrac{1}{n-3} \sum_{i=1}^{n} \hat{u}_i^2 = \dfrac{1}{n-3} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})^2$

(e) if $\text{corr}(x_1, x_2) = 0 \Rightarrow R_1^2 = 0 \Rightarrow V(\hat{\beta}_1 | x_1, x_2) = \dfrac{\hat{\sigma}^2}{\sum_{i=1}^{n} (x_{1i} - \bar{x}_1)^2}$

2. Consider a random sample of data $\{x_{1i}, x_{2i}, x_{3i}, y_i\}_{i=1}^n$ and the model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, where $E(u_i|x_{1i}, x_{2i}, x_{3i}) = 0$. With this information, answer the following questions:

(a) Suppose we wish to test $H_0 : \beta_3 = 0$. Write down the test statistic for this null.

(b) Suppose σ^2 is known, what is the distribution of the test statistic from part (a)?

(c) Suppose σ^2 is unknown, what is the distribution of the test statistic from part (a)?

(d) Suppose we wish to test $H_0 : \beta_2 = \beta_3 = 0$. Write down the test statistic for this null.

(e) Suppose σ^2 is unknown, what is the distribution of the test statistic from part (d)?

(a) $H_0 : \beta_3 = 0$

$$t = \frac{\hat{\beta}_3 - 0}{se(\hat{\beta}_3)}$$

(b) $N(0,1)$

(c) t_{n-4}

(d) $H_0 : \beta_2 = \beta_3 = 0$

$$F = \frac{(SSR_R - SSR_u)/2}{(SSR_u)/(n-4)}$$ or R^2 version

(e) $F_{2,n-4}$

3. Consider the gretl output below relating the number of cigarettes smoked per day (cigs) to the individual's level of education (educ), the price of cigarettes (cigpric), their age (age) and the square of their age (agesq) and income their (income). With the output from these two models, answer the questions on the following page:

Model 1: OLS, using observations 1–807
Dependent variable: cigs

	Coefficient	Std. Error	t-ratio	p-value
const	14.7432	6.54268	2.253	0.0245
educ	−0.376440	0.169769	−2.217	0.0269
cigpric	−0.0320155	0.101909	−0.3142	0.7535
age	−0.0413708	0.0287973	−1.437	0.1512
income	0.000117819	5.59797e–005	2.105	0.0356

Mean dependent var	8.686493	S.D. dependent var	13.72152
Sum squared resid	150157.2	S.E. of regression	13.68314
R^2	0.010520	Adjusted R^2	0.005585
$F(4, 802)$	2.131747	P-value(F)	0.075114
Log-likelihood	−3253.821	Akaike criterion	6517.641
Schwarz criterion	6541.108	Hannan–Quinn	6526.652

Model 2: OLS, using observations 1–807
Dependent variable: cigs

	Coefficient	Std. Error	t-ratio	p-value
const	1.87774	6.87287	0.2732	0.7848
educ	−0.504037	0.168659	−2.988	0.0029
cigpric	−0.0345002	0.100216	−0.3443	0.7307
age	0.796047	0.159838	4.980	0.0000
income	4.13093e–005	5.68945e–005	0.7261	0.4680
agesq	−0.00927067	0.00174150	−5.323	0.0000

Mean dependent var	8.686493	S.D. dependent var	13.72152
Sum squared resid	145026.3	S.E. of regression	13.45573
R^2	0.044331	Adjusted R^2	0.038365
$F(5, 801)$	7.431220	P-value(F)	7.94e–07
Log-likelihood	−3239.792	Akaike criterion	6491.584
Schwarz criterion	6519.744	Hannan–Quinn	6502.397

(a) Write down the marginal effect of age from model 1.

(b) Test the null hypothesis that the coefficient on age is zero in model 1.

(c) Write down the marginal effect of age in model 2.

(d) Test the null hypothesis that the number of cigarettes smoked per day is a linear function of age.

(e) Using at least two measures of goodness-of-fit, which model is preferable?

(a) $\frac{\partial y}{\partial age} = -0.0413$

(b) $H_0: \beta_{age} = 0$

$t = \frac{-0.0413 - 0}{0.0288} = -1.437 < 2$

$\Rightarrow$ fail to reject

(c) $\frac{\partial y}{\partial age} = 0.796047 + 2(-0.00927)age$

(d) $H_0: \beta_{age^2} = 0$ or F-test

$t = \frac{-0.00927 - 0}{0.00174} = |-5.323| > 2$

$\Rightarrow$ reject the null

(e) model 2: R^2, $\bar{R}^2$, $\hat{\sigma}$, log-like, AIC, SC, HQ