ECON 4230 Intermediate Econometric Theory

Solutions to Problem Set #2

Directions. **Make sure your answers are typed and tables are appropriately formatted. Turn in a hard copy at the beginning of class on the date above.**

#1. Gujarati and Porter, 5^{th} Edition, Exercise 5.1.

SOLUTION:

(a) ***True*. The t test is based on variables with a normal distribution. Since the estimators of β_1 and β_2 are linear combinations of the error u_i, which is assumed to be normally distributed under CLRM, these estimators are also normally distributed.**

(b) ***True*. So long as E(u_i) = 0, the OLS estimators are unbiased. No probabilistic assumptions are required to establish unbiasedness.**

(c) ***True*. In this case the Equation (1) in Appendix 3A, Section 3A.1, will be absent.**
***Uncertain*. By coincidence, the residuals *could* still add up to zero.**

(d) ***False*. See page 835. The size of the test is the preselected α value, such as 1, 5, or 10 percent (i.e., the level of significance of a test). The p value is defined as the lowest significance level at which a null hypothesis can be rejected.**

(e) ***True*. This follows from Equation (1) of Appendix 3A, Section 3A.1**

(f) ***False*. All we can say is that the data at hand do not permit us to reject the null hypothesis.**
***Uncertain*. It *could* be true, but we cannot definitively say that. All we can say is that the data at hand do not permit us to reject the null hypothesis.**

(g) ***False*. A larger σ^2 may be counterbalanced by a larger $\sum x_i^2$. It is only if the latter is held constant, the statement is true.**
***True*. Keeping $\sum x_i^2$ constrant, a larger σ^2 will give a lartger variance. It follows from Equation (3.3.1).**

(h) ***False*. The conditional mean of a random variable depends on the values taken by another (conditioning) variable. Only if the two variables are independent, that the conditional and unconditional means can be the same.**

(i) ***True*. This follows directly from Equation (3.1.7).**

(j) ***True*. Refer of Equation (3.5.2). If X has no influence on Y, $\hat{\beta}_2$ will be zero, in which case $\sum y_i^2 = \sum \hat{u}_i^2$.**

#2. Gujarati and Porter, 5th Edition, Exercise 5.3.

<u>SOLUTION</u>:

(a) The statistic is $t = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{14.4773}{6.1523} = 2.353$. The null hypothesis under evaluation is $H_0: \beta_1 = 0$. With $n - k = 34 - 2 = 32$ degrees of freedom and assuming a 2-sided test, the critical t value for a 5% significance level is 2.042 and we reject the null hypothesis.

(b) The statistic is $t = \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{0.0022}{0.00032} = 6.875$. In this case we use a 1-sided test, because we think the number of phones will increase as income rises. The null hypothesis under evaluation is $H_0: \beta_2 \leq 0$ versus the alternative $H_A: \beta_2 > 0$. Then the critical *t* value is 1.697 and we reject the null hypothesis.

(c) The 95% confidence interval for the true slope coefficient is

$$\hat{\beta}_2 \pm t_{0.025} se(\hat{\beta}_2)$$
$$= 0.0022 \pm 2.042 * 0.00032$$
$$= 0.0022 \pm 0.00065.$$

The slope coefficient 95% confidence interval is between 0.00155 and 0.00285.

(d) The mean forecast value is $14.4773 + 0.0022 * 9000 = 34.28$ phones per 100 persons.

For the prediction confidence interval (using equation (5.10.6)), we first need to compute $var(\hat{Y}_0) = \sigma^2[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}]$, where $\hat{\sigma}^2 = \frac{\sum \hat{u}^2}{n-k} = \frac{13{,}520.84}{32} = 422.526$ is used for σ^2. Then

$$var\left(\hat{Y}_0\right) = 422.526\left[1 + \frac{1}{34} + \frac{(9000 - 15{,}819.865)^2}{4{,}159{,}175{,}829}\right] = 439.6782.$$

The confidence interval is given as

$$Pr[\hat{Y}_0 - t_{\alpha/2} se(\hat{Y}_0) \leq Y_0 \leq \hat{Y}_0 + t_{\alpha/2} se(\hat{Y}_0)] = 1 - \alpha$$

and using $se(\hat{Y}_0) = \sqrt{439.6782} = 20.9685$.

$$Pr[34.28 - 2.042(20.9685) \leq Y_0 \leq 34.28 + 2.042(20.9685)] = 0.95.$$

The 95% confidence interval for the forecast value is between -8.54 and 77.09.

#3. Gujarati and Porter, 5th Edition, Exercise 5.17.

<u>SOLUTION</u>:

(e) **Letting *Y* represent the male math score and *X* the female math score, we obtain the following regression:**

$$\begin{aligned}\hat{Y}_t &= 198.74 + 0.67X_t \\ se &= (12.88)(0.026) \\ t &= (15.44)(25.33) \\ R^2 &= 0.95.\end{aligned}$$

(b) **The Jarque-Bera statistic is 1.164 with a *p* value of 0.414. Therefore, we cannot reject the normality assumption.**

(c) **The statistic is $t = \frac{0.67-1}{0.026} = -12.692$. Therefore, with 99% confidence we can reject the null hypothesis that $\beta_2 = 1$.**

ANOVA table:

	df	SS	F	p value
ESS	1	1605.92	641.73	0
RSS	34	85.08		

#4. Gujarati and Porter, 5th Edition, Exercise 6.1.

SOLUTION:

***True*. Note that the usual OLS formula to estimate the intercept is: $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$. But when *Y* and *X* are in deviation form, their mean values are always zero. Hence in this case the estimated intercept is also zero.**

#5. Gujarati and Porter, 5th Edition, Exercise 6.8.

SOLUTION:

The null hypothesis is that the true slope coefficient is 0.005. The alternative hypothesis is that $\beta_2 \neq 0.005$. The estimated slope value is 0.00705. Using the t-test, we obtain:

$$t = \frac{0.00705 - 0.005}{0.00018} = 11.389.$$

The critical *t* value with $n - k = 15 - 2 = 12$ degrees of freedom for a two-sided 5% significance level is 2.160. We can therefore reject the null hypothesis.

#6. Gujarati and Porter, 5th Edition, Exercise 7.12.

SOLUTION:

(a) Rewrite model B as:

$$\begin{aligned} \boldsymbol{Y_t} &= \boldsymbol{\beta_1 + (1 + \beta_2)X_{2t} + \beta_3 X_{3t} + u_t} \\ &= \boldsymbol{\beta_1 + \widetilde{\beta_2}\, X_{2t} + \beta_3 X_{3t} + u_t} \end{aligned}$$

where $\widetilde{\boldsymbol{\beta}}_{\mathbf{2}} = (\mathbf{1} + \boldsymbol{\beta_2})$. Therefore, the two models are similar. Yes, the intercepts in the model are the same.

(b) The OLS estimates of the slope coefficient of X_3 in the two models will be the same.

(c) $\widetilde{\boldsymbol{\beta}}_{\mathbf{2}} = (\mathbf{1} + \boldsymbol{\beta_2}) = \boldsymbol{\alpha_2}$.

(d) No, because the regressands in the two models are different.

#7. Download annual data from FRED and estimate a linear (X-Y) version of Okun's Law over the period 1950-2016. Then estimate a version of Okun's Law with a Box-Cox transformation on the right-hand-side variable. Comment on the results and do a *t* test to see if Okun's Law is indeed linear. [Hint: The Box-Cox methodology does not work with negative values. Since both the change in unemployment and real GDP growth are likely to be negative for some years, you may need to make the change in unemployment the "Y" variable and add a constant to all real GDP values such that no "X" values are negative.]

SOLUTION:

The Box-Cox regression results from *Stata* are shown below …

Table 1. Box-Cox Estimates of Okun's Law

Variable	Coefficient Estimates
Intercept	**4.1152**
Real GDP	**-2.093**
Box-Cox Transformation Parameter(λ)	**0.0331** **(0.2056)**

Notes. Dependent variable is annual U.S. changes in the unemployment rate. Sample period 1950 – 2016. Box-Cox transformation is only on the right-side variable. Also, because real GDP is occasionally negative, I have added 4% to all the annual real GDP observations. N = 67.

To test for a linear Okun's law model, we set up the null and alternative hypotheses as …

H$_0$**:** $\boldsymbol{\lambda = 1}$
H$_A$**:** $\boldsymbol{\lambda \neq 1.}$

For a significance level of $\boldsymbol{\alpha = 0.05}$ **and a degrees of freedom of** $\boldsymbol{n - k = 67 - 3 = 64}$**, the critical values are approximately** $\boldsymbol{\pm 2}$**. The** ***t*** **statistic (regression in levels) is**

$$\boldsymbol{t = \frac{0.0331 - 1}{0.2056} = -4.7028,}$$

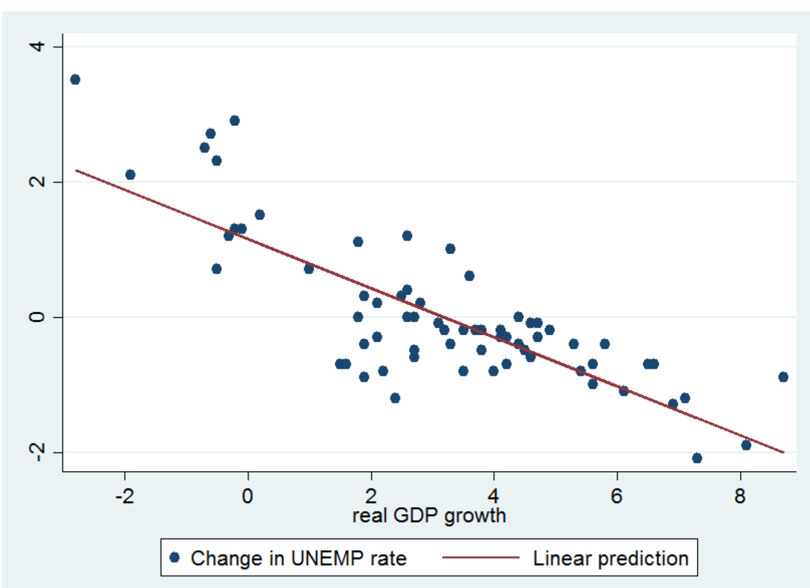
which leads to a rejection of the null hypothesis that Okun's Law is linear.

#8. Using the same data from problem #7, estimate a multivariate linear version of Okun's law which adds the labor force participation rate. Then provide both unconditional and conditional scatterplots between the growth in real GDP and changes in the unemployment rate. In each scatterplot, show the best-fitting regression line along with the slope of the line and goodness-of-fit measure. Comment on the results.

<u>SOLUTION</u>:

The unconditional scatter plot (Figure 1) is given by with a slope of -0.3635 and an $\boldsymbol{R^2 = 0.64}$**.**

Figure 1. Unconditional Scatter Plot with OLS Regression Line



Below (in Table 2) are the OLS regression results for a multivariate linear version of Okun's Law and the associated conditional (partial) regression plot:

Table 2. Multivariate OLS Estimates of Okun's Law

Variable	Coefficient Estimates
Intercept	5.7272*** (1.6832)
Real GDP	-0.3872*** (0.0337)
Labor Force Participation Rates	-0.0714*** (0.0262)
R^2	0.6738

Notes. Dependent variable is annual U.S. changes in the unemployment rate. Sample period 1950 – 2016. Box-Cox transformation is only on the right side of the regression equation. Also, because real GDP is occasionally negative, I have added 4% to all the annual real GDP observations. N = 67.
(***) indicates significant at the 1% level.

The conditional scatter plot (Figure 2) holds the labor force participation rate (LFPR) constant and is given by with a slope of -0.3872 and an $R^2 = 0.67$.

Figure 2. Conditional (on LFPR) Scatter Plot with OLS Regression Line