ECON 4230 Intermediate Econometric Theory
Solutions to the Exam

Multiple Choice (20 pts). Circle the best answer.

1. The Classical assumption of mean zero errors is satisfied if the regression model

   a) is linear in the parameters.
   b) is linear in the variables.
   c) has an R^2 greater than zero.
   d) contains a constant term.

2. The double log functional form for the regression model

   a) violates the Classical assumption of linearity.
   b) is convenient for estimating elasticities.
   c) always produces heteroscedasticity.
   d) always has the best goodness of fit.

3. The adjusted R^2

   a) is sometimes bigger than the regular R^2.
   b) is equal to the F statistic.
   c) can be used for model selection.
   d) increases with the number of regressors.

4. The p-value in hypothesis tests

   a) can never be bigger than the significance level.
   b) can never be smaller than the significance level.
   c) is equal to the significance level.
   d) can be used to reject or fail to reject a null hypothesis.

5. The probability of a Type I error is

   a) positively related to the probability of a Type II error.
   b) can be estimated with OLS.
   c) equal to the p value.
   d) equal to the significance level or size of a test.

6. The dummy variable trap

   a) only happens with spline regressions.
   b) happens when dummy variables are correlated with the errors.
   c) involves failure to omit the dummy variable for the reference category.
   d) causes autocorrelation.

7. The Box-Cox regression model

   a) violates the Classical assumption of linearity.
   b) is linear in the parameters.
   c) is nonlinear in the parameters.
   d) cannot be linear in the variables.

8. The Durbin Watson statistic

   a) has a student *t* distribution.
   b) has a standard normal distribution.
   c) is subject to Type I errors.
   d) is unrelated to the autocorrelation coefficient of the AR(1) process.

9. The slope coefficient is biased for all the reasons below except when

   a) the intercept is suppressed.
   b) there is severe multicollinearity.
   c) the error term is correlated with the explanatory variable.
   d) the model is misspecified.

10. The Central Limit Theorem

   a) is used in hypothesis testing when the errors are not normally distributed.
   b) is used in establishing the Gauss-Markov theorem.
   c) states that GLS is more efficient than OLS.
   d) can be used to test for heteroskedasticity.

#11. (20 pts) Derive the OLS estimator for a regression model with an intercept but no explanatory variables. Check the second-order conditions. Draw a graph of the population and sample regression functions with four data points. Make sure to label the axes, each line, one error term, and one residual.

Now assume you wish to test that the first two data points (say, males) have greater Y values than the last two data points (say, females). How would you do this using a dummy variable model? Explain the procedure.

SOLUTION: The regression model is $Y_i = \beta_1 + u_i$. The OLS problem is

$$\min_{\hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1)^2 .$$

The first-order condition is

$$-2\sum_{i=1}^{n}(Y_i - \hat{\beta}_1) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} Y_i - n\hat{\beta}_1 = 0 \quad \Rightarrow \hat{\beta}_1 = \bar{Y}.$$

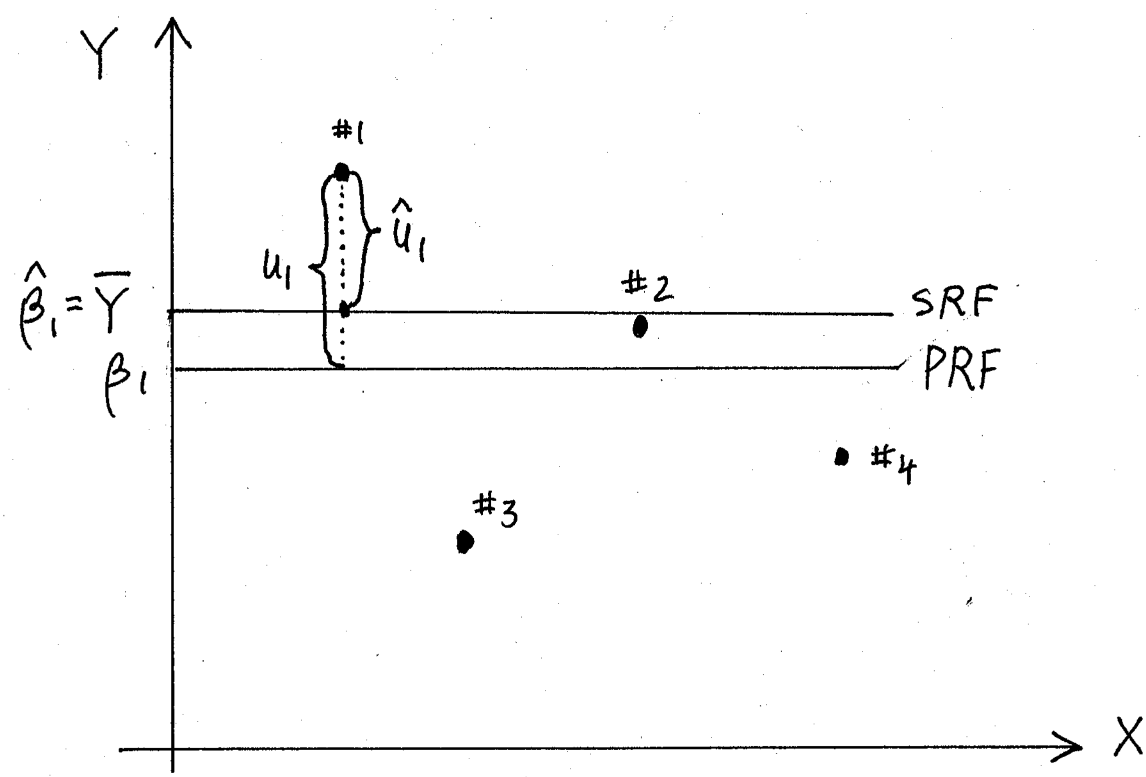
The second-order condition is

$$\frac{d^2(\sum_{i=1}^{n}\hat{u}_i^2)}{d(\hat{\beta}_1)^2} = 2n > 0.$$

Let $Male_i = 1$ and $Female_i = 1$ be the two dummy variables. To avoid the dummy variable trap, exclude one dummy variable from the model. The model is then

$$Y_i = \beta_1 + \beta_2 Male_i + u_i$$

and a one-sided *t* test with the null hypothesis H0: $\beta_2 \leq 0$ will test whether males, on average, have a greater value for Y.



#12. (30 pts) Consider the Phillips curve model

$$\pi_t = \beta_1 + \beta_2 Unemp_t + \beta_3 \pi^e_{t+1} + \beta_4 Shocks_t + u_t,$$

where π_t is inflation, $Unemp_t$ is the unemployment rate, π^e_{t+1} is expected future inflation, $Shocks_t$ is a dummy variable for the OPEC oil price shocks of the 1970s and 1980s, and $t = 1, \ldots, T$. Annual estimates for the U.S. economy between 1961 and 2012 are shown in Table 1.

a) Provide an economic interpretation of the coefficients from Table 1. Test for overall significance of the regression model.

SOLUTION:

- **The intercept is necessary but should not be interpreted.**
- **A one percentage point increase in the unemployment rate will decrease inflation by -0.38 percentage points, all else equal.**
- **A one percentage point increase in expected inflation will increase current inflation by 0.69 percentage points, all else equal.**
- **In the years with OPEC oil shocks, inflation is 3.15 percentage points higher, all else equal.**

The null hypothesis for overall significance is H_0: $\beta_2 = \beta_3 = \beta_4 = 0$. The test statistic is

$$F = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k)} = \frac{(R^2_{UR}-0)/3}{(1-R^2_{UR})/(52-3)} = \frac{(ESS/TSS)/3}{(1-(ESS/TSS))/(52-3)} = \frac{(306.51/407.02)/3}{(1-(306.51/407.02))/(52-3)} = \frac{0.753/3}{(1-0.753)/49} = \frac{0.251}{0.005} = 50.2.$$

The critical *F* statistic with 3 numerator degrees of freedom and 49 degrees of freedom is 2.8 at the 5% significance level. Therefore, we reject the null hypothesis in favor of statistical significance for the model.

b) Use a Durbin Watson statistic to test for first-order autocorrelation using both a one-tailed and two-tailed test. Make sure to clearly lay out all the steps and draw a final conclusion for both tests. Which version of the DW test do you prefer and why?

SOLUTION:

The Durbin-Watson statistic is approximately $DW = 2(1-\hat{\rho}) = 2(1-0.225) = 1.55$. The two-tailed test has null hypothesis H_0: $\rho = 0$ while the one-tailed test has null hypothesis H_0: $\rho \leq 0$. The one-tailed test has an alternative hypothesis of positive autocorrelation, which is the more likely scenario and is the more powerful test.

Using the table of critical values at the 5% significance level with $k' = 3$ and $n = 52$, the lower and upper critical values are $d_L = 1.421$ and $d_U = 1.674$. Since only one table of critical values is given for the two-tailed 5% significance level, we can say that we are inconclusive using a two-tailed test at the 5% significance level and using a one-tailed test at the 2.5% significance level.

c) Perform a hypothesis test that inflation expectations are self-fulfilling.

<u>SOLUTION</u>:

If inflation expectations are self-fulfilling, then holding all other variables constant, a change in inflation expectations will result in a one-for-one increase in inflation. This implies that $\beta_3 = 1$, which is the null hypothesis and can be tested using a *t* test.

The t statistic is

$$t = \frac{\hat{\beta}_3 - 1}{stderr(\hat{\beta}_3)} = \frac{0.6924 - 1}{0.1019} = -3.0186$$

while the critical value with $n - k = 49$ degrees of freedom is about 2.0. Therefore, we reject the null hypothesis that inflation expectations are (perfectly, one-for-one) self-fulfilling.

#13. (30 pts) Consider the following housing price hedonics regression model:

$$hprice_i = \beta_1 + \beta_2 avalue_i + \beta_3 bdrm_i + \beta_4 sqrft_i + u_i,$$

where $hprice_i$ is the price of a house, $avalue_i$ is the assessed value of the house, $bdrm_i$ is the number of bedrooms, and $sqrft_i$ is the square footage of the house. Table 2 displays the OLS regression results and Figure 1 shows the residuals plotted against square footage.

a) Provide an economic interpretation of the coefficients in Table 2 and explain whether the signs of the estimated coefficients match your intuition. If not, explain why.

<u>SOLUTION</u>: All else equal, we expect assessed value and square footage to have positive coefficients. As discussed in class, it is unclear whether the coefficient on bedrooms should be positive holding constant square footage. However, if assessed value is a perfect predictor of actual value, then square footage and the number of bedrooms would already be incorporated into the assessed value so that square footage and number of bedrooms have no additional explanatory power and the coefficients would be zero.

- **The intercept is necessary but should not be interpreted.**
- **A one dollar increase in the assessed value of the house will result in a 95-cent increase in the actual price, all else equal.**
- **One more bedroom increases the price by $11,836, all else equal.**
- **One hundred more square feet decreases the price by $480, all else equal.**

The signs make sense, except for square footage which should be either positive or zero (note however that the coefficient on square footage is not statistically different than zero).

b) Test the hypothesis that the assessed value is an unbiased predictor of the price of the house.

SOLUTION: If the assessed value is an unbiased predictor of the actual price, then a one dollar increase in the assessed value will result in exactly a one dollar increase in the actual price, all else equal. The resulting null hypothesis is H_0: $\beta_2 = 1$. This can be tested using a *t* statistic.

The t statistic is

$$t = \frac{\hat{\beta}_2 - 1}{stderr(\hat{\beta}_2)} = \frac{0.95 - 1}{0.10} = -0.5$$

while the critical value with $n - k = 84$ degrees of freedom is about 2.0. Therefore, we fail to reject the null hypothesis that the assessed value is an unbiased predictor of the actual price.

c) Comment on the residual plot in Figure 1. Does it make you question any of the Classical Assumptions? Explain how you would obtain efficient estimates of the parameters if the variance of the errors is directly proportional to square footage.

SOLUTION:

The residual plot in Figure 1 shows weak evidence of heteroscedasticity whereby the variance of the errors appears to increase slightly in the square footage of the house. Yes, I would question one of the Classical Assumptions, that is, the error variance is constant. If the error variance is directly proportional to the square footage of the house then we would have

$$var(u_i) = \sigma_i^2 = \sigma^2 sqrft_i$$

and the most efficient weight is ($w_i = 1/\sqrt{sqrft_i}$). To obtain the efficient estimates, we multiply the regression model through by w_i. The transformed (weighted) errors are

$$u_i^* = u_i/w_i,$$

which are homoscedastic. As a result, OLS provides the best linear unbiased estimates on the transformed model.

Table 1. OLS Estimates of the Annual U.S. Phillips Curve (1961-2012)

variable	coefficient	std. error
(Intercept)	3.3257	0.8019
Unemployment Rate	-0.3792	0.1368
Inflation Expectations	0.6924	0.1019
OPEC Oil Shocks	3.1476	1.0109
AR(1) coefficient ($\hat{\rho}$)	0.2251	
ANOVA Table		
ESS	306.51	
RSS	100.51	
TSS	407.02	

Table 2. OLS Housing Price Hedonic Regression (N = 88)

variable	coefficient	std. error
(Intercept)	-39.0639	21.5488
Assessed Value	0.9503	0.09800
Bedrooms	11.8356	6.5620
Square Footage	-0.0048	0.0167
ANOVA Table		
ESS	758,450	
RSS	159,405	
TSS	917,855	

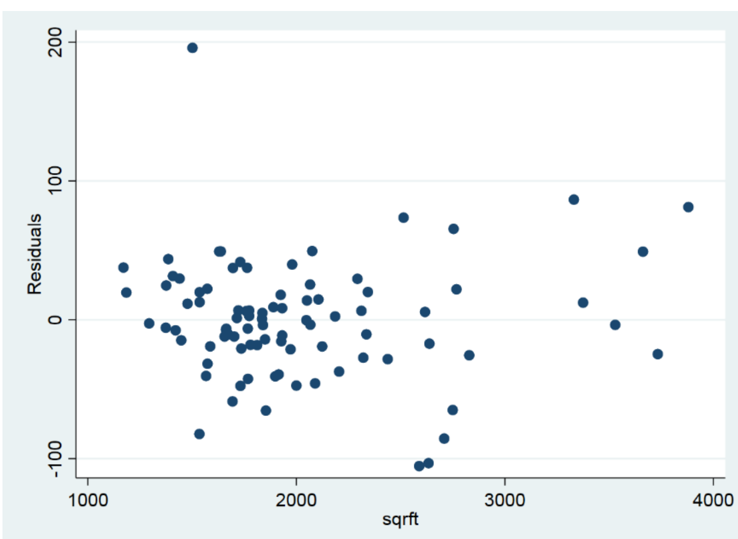


Figure 1. Housing Price Residuals vs. Square Footage