Fall 2013

ECON 5340 Applied Econometrics – Exam #2

True or False. Five points per question: 2 pts for a correct T/F and 3 pts for the explanation.

1. Measurement error in the dependent variable is more serious than measurement error in the explanatory variables.

False. Measurement error in the dependent variable only adds more noise to the model, reducing the goodness of fit, but it does not bias the coefficients. Measurement error in the independent variables, however, biases the coefficient estimates toward zero.

2. Autocorrelation biases the coefficients towards zero.

False. Autocorrelation does not bias the coefficient estimates but OLS estimates are no longer efficient.

3. Severe multicollinearity will bias the coefficient estimates.

False. Provided the multicollinearity (MC) is not perfect, MC does not bias the coefficients. OLS estimates are still unbiased and efficient. However, the standard errors are inflated relative to the case of no MC.

4. Ordinary least squares is a special case of generalized least squares.

True. Generalized least squares (GLS) involves a transformation of the variables so that the Classical assumptions hold. If there is no heteroscedasticity or autocorrelation, then OLS is simply a special case of GLS where no transformation is necessary.

5. The "runs test" for autocorrelation is immune from Type I errors.

False. A Type I error involves rejecting a true null hypothesis. This is possible if the sample is not representative and suggests autocorrelation when there is none in the data-generating process.

#6. (50 pts) Consider the earnings model: $Wage_i = \beta_1 + \beta_2 Exper_i + \beta_3 Educ_i + u_i$, where $Wage$ is the measured in dollars per hour, $Exper$ is work experience in years, and $Educ$ is the number of years of schooling. Tables 1-3 and Figure 1 show the OLS regression results for $N = 100$ males in a given year. Use the tables and figures to answer the following questions:

a) (5 pts) Using the results in Table 1, summarize the overall goodness of fit of the model. Do the signs of the coefficients match your explanations? Explain.

The overall goodness of fit is given by the R^2 value of 0.25. This implies that the 25% of the variation in the dependent variable can be explained by variation in the independent variables. The F test for overall goodness of fit is also significant. Both coefficient estimates are positive, which is to be expected.

b) (10 pts) Interpret the residual pattern in Figure 1. What conclusion do you draw? And based on that conclusion, what are the impacts on the OLS estimates in Table 1? Explain.

To my eye, I see heteroscedasticity. The variance of the errors appears to increase with Education. If this is true, the OLS estimates are unbiased but inefficient. GLS would provide efficient estimates.

c) (10 pts) Using the results in Table 3, perform White's test for heteroscedasticity. Be sure to carefully set up the null and alternative hypotheses and draw a conclusion.

White's test involves running a regression with squared residuals on the left side of the regression model and the levels, squares and crosses of all the explanatory variables on the right side. The results are shown in Table 3. The test statistic is chi squared with 5 degrees of freedom. The statistic is $nR^2 = 100 * 0.0949 = 9.49$. The critical value at the 5% level is 11.07. Therefore, we fail to reject the null of no heteroscedasticity.

d) (10 pts) Based on your answers to parts (b) and (c), describe a procedure to obtain the efficient estimators for the coefficients.

The results in parts (b) and (c) differ. However, if we assume there is some heteroscedasticity related to, say, the square of Education (i.e., $\sigma_i^2 = \sigma^2 Educ_i^2$*), then the efficient thing to do is GLS. We would then transform all the variables by dividing them by Education and running OLS on the transformed variables.*

e) (5 pts) Use the results in Table 2 to discuss the severity of the multicollinearity and the likely impacts on the OLS results in Table 1.

The pairwise correlation between Education and Experience is -0.3665. This is not strong enough to be a concern, but it will inflate the standard errors relative to the case where there is no correlation. The OLS estimates are still unbiased and efficient.

f) (10 pts) Show how to perform a *t* test that the EXPER and EDUC coefficients are equal. Do you have all the necessary information to complete the test? If so, complete the test.

The null hypothesis is $H_0: \beta_2 = \beta_3$*. The t statistic would be*

$$t = \frac{\hat{\beta}_2 - \hat{\beta}_3}{\sqrt{var(\hat{\beta}_2) + var(\hat{\beta}_3) - 2cov(\hat{\beta}_2, \hat{\beta}_3)}}.$$

The covariance between the two estimates is not provided so the test cannot be completed.

#7. (25 pts) Provide brief answers to the following three questions.

a) (5 pts) Write down an AR(1) process for the error terms and describe how to perform a Durbin-Watson test for autocorrelation.

An AR(1) process is $u_t = \rho\, u_{t-1} + \epsilon_t$. *The Durbin Watson statistic is* $DW \cong 2(1-\hat{\rho})$. *The null hypothesis is no autocorrelation. The DW statistic is calculated and compared to the low and high critical values from the DW table.*

b) (10 pts) Re-specify the earnings model in question #6, to test the hypothesis that men in a labor union earn more on average than men that are not members of a labor union. How would you test this hypothesis? Draw an appropriate figure of the regression model to support your test.

The new model would be: $Wage_i = \beta_1 + \beta_2 Exper_i + \beta_3 Educ_i + \beta_4 Union_i + u_i$, *where Union is a dummy variable for union membership. The model results in two parallel regression lines, a higher one for union men and a lower one for non-union men. The hypothesis can be tested with a t test for* $\beta_4 = 0$.

c) (10 pts) Re-specify the earnings model in question #6, to test the hypothesis that men in a labor union earn more for each additional year of experience than men that are not members of a labor union. How would you test this hypothesis? Draw an appropriate figure of the regression model to support your test.

The new model is

$$Wage_i = \beta_1 + \beta_2 Exper_i + \beta_3 Educ_i + \beta_4 Union_i + \beta_5 (Exper_i * Union_i) + u_i,$$

where Union is a dummy variable for union membership. The model results in two regression lines with different intercepts and slopes. If the theory is correct, the union membership regression line has a steeper slope. The hypothesis can be tested with a t test for $\beta_5 = 0$.

BONUS QUESTION. (5 pts) Describe an economic situation (other than the one in my lecture notes) where panel data will allow you tease out two impacts that would not be possible using only cross-section or only time-series data.

Table 1. STATA results from OLS estimation of the earnings model

Source	SS	df	MS
Model	2057.5037	2	1028.75185
Residual	6059.71269	97	62.4712648
Total	8117.21639	99	81.9920847

Number of obs = 100
F(2, 97) = 16.47
Prob > F = 0.0000
R-squared = 0.2535
Adj R-squared = 0.2381
Root MSE = 7.9039

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Educ	1.435782	.321546	4.47	0.000	.7976026	2.073962
Exper	.328525	.0658247	4.99	0.000	.1978813	.4591687
_cons	-11.91922	4.750254	-2.51	0.014	-21.34716	-2.491275

Figure 1. Wage residuals versus education

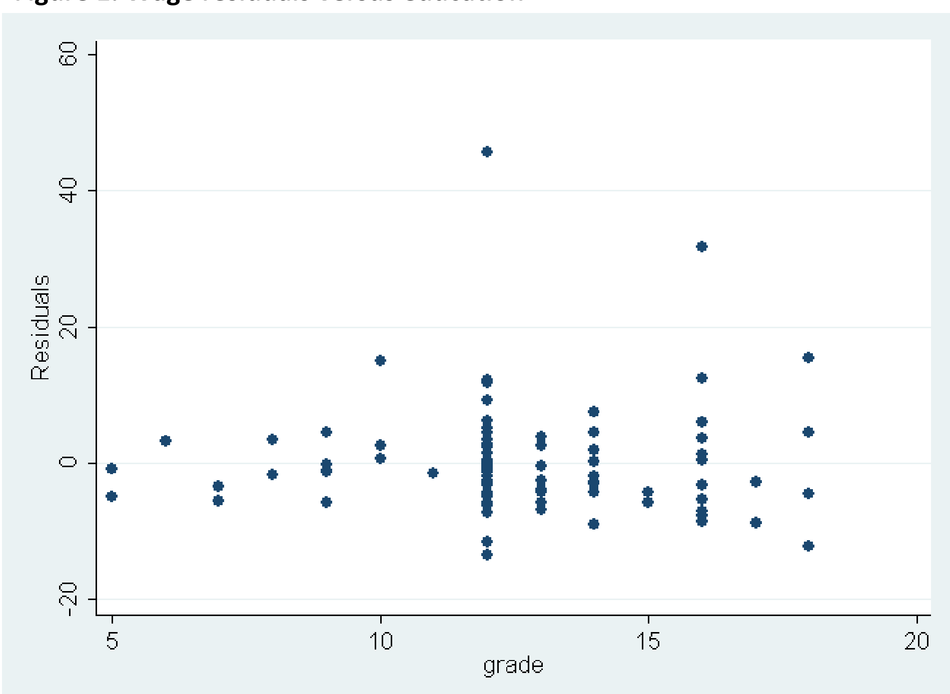

Table 2. Pairwise correlations

	grade	exper	wage
grade	1.0000		
exper	-0.3665	1.0000	
wage	0.2485	0.3163	1.0000

Table 3. STATA results with squared OLS residuals as the dependent variable

Source	SS	df	MS
Model	498933.661	5	99786.7323
Residual	4759291.93	94	50630.7652
Total	5258225.59	99	53113.3898

Number of obs = 100
F(5, 94) = 1.97
Prob > F = 0.0901
R-squared = 0.0949
Adj R-squared = 0.0467
Root MSE = 225.01

res2	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]	
Grade	-7.357599	79.35932	-0.09	0.926	-164.9274	150.2122
Exper	-23.67913	16.87954	-1.40	0.164	-57.19386	9.835591
Grade^2	-1.048003	2.223082	-0.47	0.638	-5.461984	3.365978
Exper^2	.270444	.162453	1.66	0.099	-.0521102	.5929982
Exper*Grade	.5788711	.7165818	0.81	0.421	-.8439188	2.001661
_cons	108.2517	582.867	0.19	0.853	-1049.044	1265.548