BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
FIRST SEMESTER 2022-23

ECON F241: ECONOMETRIC METHODS **MAX MARKS: 40**
DATE: 22/DEC/2022 **TIME: hr.3.00**

COMPREHENSIVE EXAMINATION (REGULAR)

Instructions: This paper consist of two parts: PART-A (Closed Book – 20 Marks) and PART-B (Open Book – 20 Marks). After completing PART-A submit and attempt PART-B (Open Book).

ID. No: **NAME:**

PART-A (CLOSED BOOK) **(20 Marks)**

For the following questions choose the correct best answer and write the corresponding letter (A or B or C or D) in the answer sheet and also put a tick against that letter. Corrections/Overwriting/illegible answers will carry no weightage. (28 x1/2 = 14 Marks)

1) You specify a simple classical linear regression model to analyze the relationship between the dependent variable health and the explanatory variable income. Health economics studies find evidence that individuals with more education tend to have better health. Labor economics studies find that individuals with more education tend to have higher incomes. Given this information, which of the following would you expect?

A. The error term will be positively correlated with income.
B. The error term will be negatively correlated with income.
C. The error term will have non-constant variance.
D. The errors for the different households will be correlated.

2) According to the Central Limit Theorem:

A. the OLS estimators are BLUE at all sample sizes.
B. the OLS estimators are BLUE at sufficiently large sample sizes.
C. the OLS estimators are normally distributed at sufficiently large sample sizes.
D. the OLS estimators are unbiased at sufficiently large sample sizes.

3) The simple classical linear regression model makes the assumption that that error term has mean zero. This implies:

A. There errors for any two units in the sample are uncorrelated.
B. The error term and the explanatory variable are uncorrelated.
C. The error term has a constant variance.
D. The error term has a normal distribution.

4) The assumption that the error term is normally distributed: ui ~ $N(0, \sigma^2)$ is required in order for the least-squares estimators to be

A. consistent.
B. best linear unbiased estimators.
C. best unbiased estimators.
D. Asymptotically normally-distributed (that is, approximately normally-distributed in large samples).

5) According to the following model: $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i2}^2$, a one-unit increase in x_{i2} will cause y_i to increase by about

A. $(\beta_1 + \beta_2)$ units.
B. $(\beta_2 + \beta_3)$ units.
C. $(\beta_2 + \beta_3 x_{i2})$ units.
D. $(\beta_2 + 2\beta_3 x_{i2})$ units.

6) Suppose Q = quantity demanded, P = price of the good, and I = consumer income. In which specification does β_2 equal the price elasticity of demand?

A. $Q_i = \beta_1 + \beta_2 P_i + \beta_3 I_i$
B. $Q_i = \beta_1 + \beta_2 (P_i/I)$
C. $Q_i = \beta_1 + \beta_2 \ln(P_i) + \beta_3 \ln(I_i)$
D. $\ln(Q_i) = \beta_1 + \beta_2 \ln(P_i) + \beta_3 \ln(I_i)$

7) Suppose X_1 and X_2 are two of the independent variables in a linear regression model and further suppose that, $X_{1i} = 0{:}75 + X_{2i}$; for all i =1,..., n. This is a violation of which "Classical Assumption?"

A. The stochastic error terms are serially uncorrelated across observations.
B. The stochastic error terms are serially correlated across observations
C. No explanatory variable is a perfect linear function of any other explanatory variable.
D. No explanatory variable is a perfect nonlinear function of any other explanatory variable.

8) Consider two different linear estimators, $\hat{\beta}$ and $\tilde{\beta}$, of a population parameter β from a linear regression model. Suppose $E(\hat{\beta}) = \beta$, $E(\tilde{\beta}) \neq \beta$, and $Var(\hat{\beta}) < Var(\tilde{\beta})$. Then, all else equal

A. $\hat{\beta}$ is an unambiguously better estimator than $\tilde{\beta}$.
B. $\tilde{\beta}$ is an unambiguously better estimator than $\hat{\beta}$.
C. $\tilde{\beta}$ is an unbiased estimator.
D. $\hat{\beta}$ is definitely not the OLS estimator of β.

9) Consider the following regression model:
$\log(y_i) = \beta_0 + \beta_1 \cdot \log(x_{1i}) + \beta_2 \cdot \log(x_{2i}) + u_i, i = 1,...,n.$

A. This equation is linear in the variables but not in the coefficients.
B. This equation is linear in both the variables and the coefficients.
C. This equation is linear in both the coefficients and the stochastic error term.
D. This equations parameters cannot be estimated by OLS.

10) Recall the model of first-order serial correlation: $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$, where $\rho \neq 0$ and u_t is a classical error term. Suppose $\varepsilon_{t-1} = 0.67$, $u_{t-1} = 0.23$, $\rho = 0.43$, and $u_t = 0$. Then:

A. $\varepsilon_t = 0.8821$.
B. $\varepsilon_t = 0.3388$.
C. $\varepsilon_t = 0.1228$.
D. $\varepsilon_t = 0.2881$.

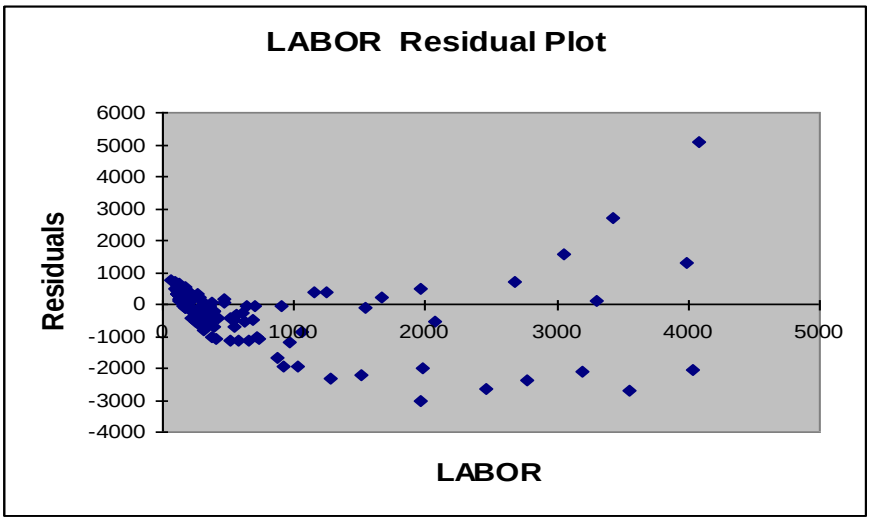
11) Which of the following is an assumption of simple linear regression?

A. a simple random sample was taken
B. For any value of X, the values of all Y are normally distributed
C. the average value of all Y is a straight line function of X
D. all of the other answers are correctly stated assumptions.

12) Which of the following is a difference between simple linear regression (SLR) and multiple linear regression (MLR)?

A. SLR uses a residual plot to check the assumptions but MLR does not.
B. SLR uses only a t-test while MLR uses both t-tests and an F-test.
C. SLR has to add the phrase "holding all other variables constant" to the interpretation of a slope.
D. R^2 is always higher in SLR

13) Based on the following plot what is a possible remedy to observed violation(s)?
A. remove multicollinearity
B. transform the dependent and possibly the independent also
C. delete the outliers
D. transform just the independent variable



14) You are fitting a dummy variable model where the independent variable is a qualitative variable with three levels. If the mean of the first level is 5, the mean of the second level is 10 and the mean of the third level is 8 and the two dummy variables are defined as (X_1 = 1 if level 1, 0 otherwise) and (X_2=1 if level 2, 0 otherwise), then which of the following is the dummy variable model?
A. $8\ X_0 - 5X_1 + 10X_2$
B. $8 - 3X_1 + 2X_2$
C. $8 - 5X_1 + 10X_2$
D. $8 + 5X_1 + 10X_2$

B- the three means are 5, 10, 8 with base level being level 3. Therefore, you start with a value of 8. To get to the first level you need to subtract 3 from the base. To get to the second level you need to add two to the base. The model is then 8-3X1 + 2X2

15) Heteroscedasticity occurs when
A. there are larger values on X than Y.
B. there is a linear relationship between X and Y.
C. more error is accounted for than remains.
D. variability in Y depends on the exact value of X.

16) When an irrelevant variable is added to a regression model:
A. the OLS estimators are no longer BLUE.
B. the OLS estimators are BLUE
C. the standard errors of the OLS estimators increase.
D. the stochastic error term exhibits serial correlation.

17) After estimating a linear regression, you obtain the residuals and the fitted values for the y's. You then regress the square of the residuals on a constant, the fitted values and the square of the fitted values. You are testing for

A. Multicollinearity
B. autocorrelation in the residuals
C. heteroscedasticity
D. goodness of fit.

18) Recall the model of first-order serial correlation: $\varepsilon_t = \rho\,\varepsilon_{t-1} + u_t$,

where $\rho \neq 0$ and u_t is a classical error term. Suppose $\varepsilon_{t-1} = 0.67$, $u_{t-1} = 0.23$, $\rho = 0.43$, and $u_t = 0$. Then:

A. $\varepsilon_t = 0.8821$.
B. $\varepsilon_t = 0.3388$.
C. $\varepsilon_t = 0.1228$.
D. $\varepsilon_t = 0.2881$.

19) suppose the true regression model is:

$$QU_t = \beta_0 + \beta_1 \cdot PU_t + \beta_2 \cdot PR_t + \varepsilon_t,$$

where QU_t is the quantity demanded of umbrellas in observation t, PU_t is the price of umbrellas in observation t, and ε_t is the stochastic error term in observation t. Suppose, however, the following model is estimated:

$$\hat{Q}U_t = \hat{\beta}_0 + \hat{\beta}_1 \cdot PU_t + \hat{\beta}_2 \cdot PR_t + \hat{\beta}_3 \cdot DMON_t,$$

where $DMON_t$ is a dummy variable which takes on the value 1 if observation t falls on a Monday and takes on the value 0 otherwise. Estimation of this misspecified model will cause:

A. positive bias in both $\hat{\beta}_1$ and $\hat{\beta}_2$.
B. negative bias in both $\hat{\beta}_1$ and $\hat{\beta}_2$.
C. unambiguous bias in $\hat{\beta}_1$ and $\hat{\beta}_2$, i.e., there will be bias, but we can't determine the direction of the bias.
D. no bias in $\hat{\beta}_1$, $\hat{\beta}_2$, or $\hat{\beta}_3$.

20) Suppose the Durbin-Watson d statistic equals 0.98, the number of independent variables is 5, a constant term is included in the model, and the sample size is 30. At the 5% significance level, the outcome of testing: $H_0 : \rho \leq 0$, against $H_A : \rho > 0$ is that:

A. the null hypothesis is not rejected.
B. the test is inconclusive.
C. the null hypothesis is rejected.
D. the alternative hypothesis is rejected.

21) In the log –log linear regression model, the regression slope:

A. indicates by how many percent Y increases, given a one percent increase in X
B. when multiplied with the explanatory variable will give you the predicted Y
C. indicates by how many unit Y increases, given a one unit increase in X
D. represents the elasticity of Y on X.

22) Consider a regression with time series data measured at quarterly intervals and let three seasonal dummy variables be defined as follows:

$$X_{1t} = \begin{cases} 1, \text{if observation t is in the first quarter} \\ 0, \text{otherwise} \end{cases}$$

$$X_{2t} = \begin{cases} 1, \text{if observation t is in the second quarter} \\ 0, \text{otherwise} \end{cases}$$

$$X_{3t} = \begin{cases} 1, \text{if observation t is in the third quarter} \\ 0, \text{otherwise} \end{cases}$$

Next consider the following estimated regression model:

$$\hat{S}_t = 15{,}600 - 4{,}500 \cdot X_{1t} + 300 \cdot X_{2t} + 62{,}500 \cdot X_{3t},\ t = 1,...,n, \text{ where}$$

S_t is the dollar amount of sales at a souvenir shop at a beach resort in period t. According to this model:

A. the expected sales in the first quarter is $15,600.
B. the expected sales in the first quarter is $300.
C. the expected sales in the first quarter is $63,500.
D. the expected sales in the first quarter is $11,100.

23) Consider the following estimated regression equation:

$$\hat{Y} = 0.32 + 1.54 \cdot X_1 + 2.33 \cdot X_2 - 1.22 \cdot X_3,$$
$$(0.11)\quad (0.26)\quad (0.45)\quad (1.01)$$

where the numbers in parentheses are standard errors. Suppose that the sample size is 22 and consider the following hypothesis test:

$$H_0 : \beta_1 = 1.0$$

$$H_A : \beta_1 \neq 1.0$$

If the significance level of the test were set to 5%, the critical t-value would be:

A. 2.120
B. 2.131
C. 2.101
D. 1.734

24) In the multiple regression model, the adjusted R2,

A. cannot be negative
B. will never be greater than, the regression R^2
C. equals the square of the correlation coefficient r
D. cannot decrease when an additional explanatory variable is added.

25) The main difference between the White and the Breusch-Pagan tests for heteroscedasticity is

A. the White test but not the Bresuch-Pagan includes the independent variables from the original equation in the test equation
B. the White test must be based on an F-Statistics while the Bresuch-Pagan test must be based on chi-squared statistics
C. the Bresuch-Pagan test includes quadratic terms while the White test includes only linear terms
D. the White test includes quadratic terms while the Bresuch-Pagan includes only linear terms.

26) When the estimated slope coefficient in the simple regression model, estimated b1 is zero then

A. R^2 = Ybar
B. $0 < R^2 < 1$
C. $R^2 = 0$
D. $R^2 >$ (SSR/TSS)

27) If you are rejecting a joint null hypothesis using the F-test in a multiple hypothesis setting, then

A. a series of t-tests may or may not give you the same conclusion
B. the regression is always significant
C. all of the hypothesis are always simultaneously rejected.
D. The F-statistics must be negative.

28) Consider the following regression model:

$$\log(y_i) = \beta_0 + \beta_1 \cdot \log(x_{1i}) + \beta_2 \cdot \log(x_{2i}) + \varepsilon_i, i = 1,...,n.$$

A. This equation is linear in the variables but not in the coefficients.
B. This equation is linear in both the variables and the coefficients.
C. This equation is linear in both the coefficients and the stochastic error term.
D. This equations parameters cannot be estimated by OLS.

Assess the following statements as TRUE or FALSE. Give a short explanation or qualification. If a statement is not true in general, but is true under some conditions, state the conditions clearly.
No credit without proper justification. (6 x 1.0 = 6 Marks)

29) Heteroscedasticity leads to an unbiased estimator of β1 cap and hence the t-test based on β1 cap and s.e. of (β_1 cap) and is valid.

30) Serial correlation leads to an unbiased estimator of β_1 cap but the t-test based on β_1 cap and s.e. of (β_1 cap) is always invalid.

31) Misspecification of the functional form leads to an unbiased estimator of β1 cap but the t-test based on β1 cap and s.e. of (β1 cap) and is invalid

32) Since even a high degree of multicollinearity (but not perfect multicollinearity) does violate the conditions of the Gauss-Markov theorem, it can be safely ignored.

33) Dummy variables are essentially a means for accounting for left out or unmeasurable variables; so their estimated coefficients have no precise meaning.

34) If X_2 is correlated with X_3, then dropping X_3 from the model $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$ will alter the OLS estimate of β_2.

Bonus Question:

I have the estimated slope coefficients from the regression of Y on X_2 and from the regression of Y on X_3. I also have the slope coefficient from the regression of X_2 on X_3, so I can compute the slope coefficients that would result from the multiple regression of Y on X_2 and X_3.

NOTE: Open Book examination. Answer to the point. Irrelevant answers will be penalised. Highlight the final answer. Write the assumptions if any, clearly.

1. Because influenza is a serious threat to the elderly population, the Indian Institute of Health and Family Welfare (IIHFW) has been stepping up its efforts induce the population over the age 75 to get vaccinated each season.
   The (IIHFW) has collected annual data over the past 20 years on the number of elderly getting the vaccine at the start of the season (Y) and the advertising budget (X). The advertising budget is in millions of rupees and the numbers vaccinated are in tens of thousands. You have the following sample statistics:

   $\sum X_t = 303$ $\sum X_t^2 = 5924$ RSS = 656

   $\sum Y_t = 1394$ $\sum Y_t^2 = 106{,}302$ $\sum X_t Y_t = 24{,}473$

a. Write the general linear form of the population regression function (PRF) and of the sample regression function (SRF), explaining how they differ in the abstract.
b. Compute the sample regression slope coefficient ($\hat{\beta}_2$) and give its interpretation (using the units given above).
c. Compute the standard error of the estimate of the slope parameter, se($\hat{\beta}_2$). What are its units and what does it mean?
d. Compute the coefficient of determination (R^2) for this regression and explain what it means.

2. An econometrician run the multiple regression and got the following output:

Dependent Variable: S
Method: Least Squares
Date: 02/28/02 Time: 10:04
Sample: 1 38
Included observations: 38

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.223923	1.935467	4.765736	0.0000
T	-0.786568	0.590256	-1.332588	0.1918
E	7.906074	3.660143	2.160045	0.0381
P	0.000408	0.000495	0.823516	0.4161
H	-0.018971	0.002674	-7.094425	0.0000

R-squared	0.709460	Mean dependent var	8.438421
Adjusted R-squared	0.674243	S.D. dependent var	2.621777
S.E. of regression	1.496382	Akaike info criterion	3.766057
Sum squared resid	73.89229	Schwarz criterion	3.981529
Log likelihood	-66.55509	F-statistic	20.14541
Durbin-Watson stat	2.146911	Prob(F-statistic)	0.000000

a) Your friend expects the coefficient of P to be positive. Carry out the formal hypothesis statement, test the hypothesis by using the 5% level of significance, and make your conclusion.
b) Your friend also argues that for every car at least should have 8.5 units of drag for safety in

the road. Construct the confidence interval for the coefficient of E, carry out the hypothesis statement and test by using the 1% level of significance and make your conclusion. The critical value of the test statistic at 0.01, 33df is **-2.457.**

c) Explain why you need to use the F test. What is your conclusion from the F hypothesis test? The critical value of $\mathbf{F^c_{(0.05,\,4,\,33)}}$ **=2.69.**

3. Consider the following model:

$$\mathbf{GNPt = \beta_1 + \beta_2 M_t + \beta_3 M_{t-1} + \beta_4 (M_t - M_{t-1}) + u_t}$$

Where GNPt = GNP at time t, M_t = money supply at time t, M_{t-1} = money supply at time (t−1), and
($M_t - M_{t-1}$) = change in the money supply between time t and time (t − 1).

This model thus postulates that the level of GNP at time t is a function of the money supply at time t and time (t- 1) as well as the change in the money supply between these time periods.

a. Assuming you have the data to estimate the preceding model, would you succeed in estimating all the coefficients of this model? Why or why not?

b. If not, what coefficients can be estimated?

c. Suppose that the $\beta_3 M_{t-1}$ term were absent from the model. Would your answer to (a) be the same?

Suppose in the model

$$\mathbf{Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i}$$

The coefficient of correlation r_{23}, between X_2 and X_3, is zero. Therefore, someone suggests that you run the following regressions:

$$\mathbf{Y_i = \alpha 1 + \alpha_2 X_{2i} + u_{1i}}$$
$$\mathbf{Y_i = \gamma_1 + \gamma_3 X_{3i} + u_{2i}}$$

d. Will α_2 cap = β_2 cap and γ_3 cap= β_3 cap? Why?

e. Derive a relation between β_1 cap and α_1 cap or γ_1 cap?

f. Will var (β_2 cap) = var (α_2 cap)?

4. The following model was estimated with time series data using the Ordinary Least Squares (OLS) procedure (with OLS standard errors reported in parentheses):

$$y_t = \underset{(0.70)}{-6.29} + \underset{(0.07)}{1.45x_t} + \hat{u}_t$$

where y and x are expressed in natural logarithms. The sequence of fifteen residuals, $\hat{u}_t$, obtained from the regression model, ranked from 1990 through to 2004, is given by:

Time (T)	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
$\hat{u}_t$	+ 0.016	+ 0.041	- 0.029	- 0.017	- 0.071	+ 0.019	- 0.001	- 0.005	+ 0.002	+ 0.011	+ 0.058	+ 0.002	+ 0.003	+ 0.021	- 0.009

where T is the observation year and $\hat{u}_t$ is the estimated residual for time period t.

From the above outline the possible sources and consequences of autocorrelated errors in a linear regression model.
Use the reported information to implement a parametric test for autocorrelated errors in this regression model. Use a significance level of 0.05 and state clearly the null and alternatives under test. Draw the inference.

5.

A. Suppose we wish to test for heteroscedasticity in the regression equation

$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 (x_1 x_2) + \varepsilon$$

using White's test. Note that in this equation, the third regressor is an interaction of the first two regressors. Recall that White's test involves estimating an auxiliary regression equation.

a. What variable would be on the left side of this auxiliary regression? That is, what would be the dependent variable?

b. What variables would be on the right side of this auxiliary regression? That is, what would be the regressors?

c. Suppose this auxiliary regression was estimated with an R^2 value of 0.06. The sample contained 200 observations. Test the null hypothesis of homoscedasticity against the alternative hypothesis of heteroscedasticity at 5 percent significance. Give the value of the test statistic, its distribution under the null hypothesis, the critical point(s), and your conclusion (accept or reject the null hypothesis).

B. The regression equation $y = \beta_1 + \beta_2 x + \varepsilon$ was estimated using 80 cross-sectional observations on countries, by ordinary least squares. To check for heteroscedasticity related to population, separate regressions were run for the 32 countries with the lowest populations and the 32 countries with the highest populations. The sum of squared residuals for the low-population countries was 240. The sum of squared residuals for the high-population countries was 90.

a. Compute unbiased estimates of the variance of the error term in the two subsamples.

b. Test the null hypothesis of homoscedasticity, against the (one-sided) alternative hypothesis that low-population countries have higher error variance, at 5 percent significance using a Goldfeld-Quandt test. Give the value of the test statistic, the critical point, and your conclusion (accept or reject the null hypothesis of homoscedasticity).

c. Suppose you also wish to test for heteroscedasticity related to population using the Breusch-Pagan test at 5 percent significance. After estimating the original regression on the entire sample of 80 countries, you saved the residuals and estimated the following auxiliary regression.

$$\hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 pop_i + \nu_i$$

where the dependent variable is the squared residual from the original equation, pop_i is population of country i and ν_i is a new error term for the auxiliary regression. This auxiliary regression produced an R^2 value of 0.05. Give the value of the test statistic, the critical point, and your conclusion (accept or reject the null hypothesis of homoskedasticity).

d. Regardless of your answer to parts (c) and (d), suppose you believe that heteroscedasticity is indeed present and that the variance of the error term is inversely proportional to population: $Var(\varepsilon_i) = \alpha/pop_i$, where α = an unknown parameter. The first observation in the raw data is shown below. Compute the transformed first observation of the data. Place your answers in the table below.

Observation (i)	x_i	y_i	pop_i	*Transformed* x_i	*Transformed* y_i
1	40	20	16		

6. The following equation describes workers' earnings:

$$W = \beta_1 + \beta_2 Age + \beta_3 Age^2 + \beta_4 Male + \beta_5 Dropout + \beta_6 College + \beta_7 Master + u$$

where W is annual earnings (in rupees), *Age* is the worker's age (in years), *Male*, *Dropout*, *College*, and *Masters* are dummy variables, and u is a residual.

The dummy variables are defined as: *Male=1* if the worker is male; 0 otherwise.
Dropout = 1 if the worker has 0-11 years of schooling; 0 otherwise.
College = 1 if the worker has 16-17 years of schooling; 0 otherwise.
Masters =1 if the worker has 18+ years of schooling; 0 otherwise.
In addition, the data set contains a dummy for high school graduates without a completed college degree,
HSgrad = 1 if the worker has 12 - 15 years of schooling; 0 otherwise.

The estimated equation over a cross-section sample of 560 workers is:

$$\widehat{W} = -16000 + 2346Age - 23Age^2 + 12000Male - 10000Dropout + 16000College + 46000Master$$

A) What is the predicted annual earnings of a 30 year old man who went to college but not to graduate school?
B) If we replace *Dropout* with *Hsgrad*, what would the coefficient estimates be?
C) If we have both *Dropout* and *Hsgrad* in the regression, what would we find?
D) Someone states: "If *Age* increases, so does Age^2. *Age* and Age^2 are thus strongly positively correlated, and the OLS estimates are biased, inconsistent, and inefficient." Comment.
E) This specification implies that the effects of schooling on wage earnings are the same by sex. Explain in detail how you would explore whether returns to schooling differ by sex. State how you would re-specify this equation and the specific null hypothesis in terms of that equation. Provide the test statistic and its degrees of freedom.

(5.0)

****