Econ 3180 Final Exam Answer Key (Winter 2015)

Part A – Multiple Choice

1. D
2. A
3. B
4. B
5. D
6. B
7. A
8. D
9. C
10. A
11. D
12. A
13. B
14. C
15. D
16. B

Part B - Short Answer

1) Viewing a scatterplot can reveal several important features of the data, such as:

- whether there appears to be association between two variables, either positive or negative
- whether the association between two variables is linear or non-linear
- whether there are any outliers
- whether the data is characteristic of homoskedasticity or heteroskedasticity

2) If the error terms are assumed to be homoskedastic when they are actually heteroskedastic, then the formula for the $var(\hat{\beta})$ is wrong. Estimation of $var(\hat{\beta})$ will be inconsistent. t-statistics ~~will~~ and p-values will be wrong, and hypothesis testing in general will be invalid.

3) $R^2 = ESS/TSS$. R^2 has the property that it must increase (or at best stay the same) when any variable is added to the model, even if the variable isn't relavent. This is because ESS can't decrease with the addition of a variable (OLS has more "options"). Adjusted R-square, however, penalizes the model for the number of regressors, K, and is a more appropriate measure of fit since it can increase or decrease with the addition of a varieble.

4) The t-test can't be used. The null hypothesis is wrong if $\beta_2 \neq 0$, if $\beta_3 \neq 0$, or if β_2 and β_3 are both not equal to zero. Even if the individual t-stats are not correlated, the test will not have the right type I error (significance). This can be corrected by choosing the right critical value. However, in practice, X_2 and X_3 will exhibit some sample correlation, and so the t-stats will be correlated. This correlation needs to be taken into account, and is done so by the F-test.

5)

6) u_i comes from the population model. u_i is unobservable, and contain omitted/unobservable factors which influence the dependent variable, y.

$\hat{u}_i$ comes from the fitted model. $\hat{u}_i$ is observable, and is the "prediction error" of the fitted model.

7) An unrestricted model is one in which all of the parameters (the β^s) are free to be estimated (by OLS for example). A restricted model is one in which some of the parameters are chosen, for example by a null hypothesis. An unrestricted model might be:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$$

If the null hypothesis is $\beta_2 = 0$, the restricted model implied by this null is:

$$y = \beta_0 + \beta_1 X_1 + \varepsilon$$

8) We can either take the derivative of TestScore with respect to income, or we can consider a specific change in income. Since we have a non-linear model, the effect will depend on the value of income itself.

Derivative: $\dfrac{\partial \widehat{TestScore}}{\partial Income} = 3.85 - 0.0846\,\underline{Income}$

Change in $\widehat{TestScore}$ due to a change in Income from 10,000 to 11,000:

$$\Delta \widehat{TestScore} = 3.85(1000) - 0.0423(11000^2) + 0.0423(10000^2)$$

9) The OLS estimates are:

$$\hat{\beta}_0 = 14.88, \quad \hat{\beta}_1 = 7.10, \quad \hat{\beta}_2 = -2.96, \quad \hat{\beta}_3 = -0.55$$

10) The estimated effect has changed significantly, likely due to omitted variable bias.

Living.Area is likely an important determinant of house price. Fireplaces is likely correlated with Living.Area, as a high number of fireplaces are found in larger houses, etc. In the first model, the omitted variable is (i) a determinant of the dependent variable and (ii) correlated with the included variable. This will cause O.V.B.

11) If the effect of X on Y is non-linear, but a linear model is specified, then OLS is likely biased and inconsistent. So it is important to capture non-linear effects in this respect.

A non-linear effect means that the effect that X has on Y depends on the value of X.

Non-linear effects can be captured using interaction terms, polynomials, and logarithms.

Part C - Long Answer

a) $R^2 = 1 - \frac{SSR}{TSS}$

$\bar{R}^2 = 1 - \frac{SSR}{TSS}\left(\frac{n-1}{n-k-1}\right)$

Since $\frac{n-1}{n-k-1} > 1$, $R^2 > \bar{R}^2$ (for $k > 0$).

b) R^2 can not be used to compare models (1) and (2), since the dependent variables are different. R^2 measures the proportion of variation in the dependent variable that can be explained by variation in the independent variables. Once we take the log of AHE, we change the variance of the dependent variable.

c) Even though the intercept appears statistically insignificant, it should not be dropped from the model. The inclusion of the intercept is mostly for algebraic convenience, and in most econometric models, holds no meaningful interpretation. ~~For~~ In this model, the intercept would by the ln(AHE) for a male who doesn't have a B.A., and who is Ø years old. It doesn't make economic sense to interpret the intercept in this model.

d) This estimated coefficient may be interpreted as:

a 1% increase in Age is associated with a 0.725% increase in AHE.

e) $\ln(\widehat{AHE}) = 0.147(25) - 0.002(25^2) + 0.405 + 0.059$
$= \cancel{5.389}\ 2.889$

f) Models (4) - (8) may be used. In each case, the null hypothesis of linearity is rejected.

g) In models (5) - (8), we can see that the "Female×Bachelor" variable is significant at the 1% level. Hence, the effect of a bachelor's degree appears to be different for women than for men.

h) We can use model (6) or (8). We must jointly test whether the coefficients on "Female×Age" and "Female×Age^2" are equal to zero. This can be accomplished using the R^2 from models (5) and (6), or (7) and (8).

i) $\ln(\widehat{AHE})\,|_{Female,\, Age=25,\, BA=1} = 0.16(25) - 0.002(25^2)$
$- 0.123(25) + 0.002(25^2) + 0.091(25) - 0.001(25^2)$
$+ 1.764 - 1.186 + 0.066 =$

.53

$\ln(\widehat{AHE})\,|_{Female,\, Age=25,\, BA=0} = 0.16$

.464