Prof. C. M. Dalton
ECN 209A
Spring 2015
Practice Problems (After HW1, HW2, before HW3)

Question 1. Draw and describe a relationship with heteroskedastic errors. Support your claim with a brief explanation of the relationship between two variables and draw a representative scatterplot.

Question 2.

a) List and define all assumptions for multiple OLS regression.

b) Do heteroskedastic errors violate any of the above assumptions? Explain.

Question 3.

a) Define adjusted R-squared. (formula)

b) What does adjusted R-squared account for compared with standard R-squared? E.g. Why do we bother adjusting?

Question 4. The cost of attending your college has once again gone up. Although you have been told that education is investment in human capital, which carries a return of roughly 10% a year, you (and your parents) are not pleased. One of the administrators at your university does not make the situation better by telling you that you pay more because the reputation of your institution is better than that of others. To investigate this hypothesis, you collect data randomly for 100 national universities and liberal arts colleges from the 2000-2001 *U.S. News and World Report* annual rankings. Next you perform the following regression

$$\underset{}{\textit{Cost}} = \underset{(580.2)}{7{,}311.17} + \underset{(524.4)}{3{,}985.20} \times \textit{Reputation} - \underset{(0.03)}{0.20} \times \textit{Size} + \underset{(3502.7)}{8{,}406.79} \times \textit{Dpriv} - \underset{(204.1)}{416.38} \times \textit{Dlibart} - \underset{(1200.0)}{2{,}376.51} \times \textit{Dreligion}$$

R^2=0.72, *SER* = 3,773.35 and coefficient standard errors are in ().

where *Cost* is Tuition, Fees, Room and Board in dollars,
Reputation is the index used in *U.S. News and World Report* (based on a survey of university presidents and chief academic officers), which ranges from 1 ("marginal") to 5 ("distinguished"),
Size is the number of undergraduate students, and
Dpriv, *Dlibart*, and *Dreligion* are binary variables indicating whether the institution is private, a liberal arts college, and has a religious affiliation.

a. Interpret the results. (Remember what "interpret" means?!) Do the coefficients have the expected sign?

b. What is the forecasted cost for a liberal arts college, which has no religious affiliation, a size of 1,500 students and a reputation level of 4.5? (All liberal arts colleges are private.)

c. To save money, you are willing to switch from a private university to a public university, which has a ranking of 0.5 less and 10,000 more students. What is the effect on your cost? Is it substantial (support this)?

d. Do you have a reason to suspect imperfect multicollinearity in the independent variables above? Why or why not? Describe what happens to an independent variable's estimated coefficient if imperfect multicollinearity is present.

e. Eliminating the *Size* and *Dlibart* variables from your regression, the estimation regression becomes

Cost = 5,450.35 + 3,538.84 × *Reputation* + 10,935.70 × *Dpriv* – 2,783.31 × *Dreligion*;
R^2=0.68, *SER* = 3,792.68

Why do you think that the effect of attending a private institution has increased now?

f. Describe a variable that you could put into the regression (but won't!) that would be perfectly collinear with another variable(s) in the original regression. Also, why won't you add it?!

These questions deal with Chpt 8, Nonlinear Regression

Question 5 Sports economics typically looks at winning percentages of sports teams as one of various outputs, and estimates production functions by analyzing the relationship between the winning percentage and inputs. In Major League Baseball (MLB), the determinants of winning are quality pitching and batting. All 30 MLB teams for the 1999 season. Pitching quality is approximated by "Team Earned Run Average" (ERA), and hitting quality by "On Base Plus Slugging Percentage" (OPS).

Summary of the Distribution of Winning Percentage, On Base Plus Slugging Percentage, and Team Earned Run Average for MLB in 1999

	Average	**Standard deviation**	**Percentile**						
			10%	**25%**	**40%**	**50% (median)**	**60%**	**75%**	**90%**
Team ERA	4.71	0.53	3.84	4.35	4.72	4.78	4.91	5.06	5.25
OPS	0.778	0.034	0.720	0.754	0.769	0.780	0.790	0.798	0.820
Winning Percentage	0.50	0.08	0.40	0.43	0.46	0.48	0.49	0.59	0.60

Your regression output is:

$$\widehat{Winpct} = -0.19 - 0.099 \times teamera + 1.490 \times ops$$
$$(0.08)\ (0.008) \qquad (0.126)$$

R^2=0.92, *SER* = 0.02.

(a) Interpret the regression. Are the results statistically significant and important?

(b) There are two leagues in MLB, the American League (AL) and the National League (NL). One major difference is that the pitcher in the AL does not have to bat. Instead there is a "designated hitter" in the hitting line-up. You are concerned that, as a result, there is a different effect of pitching and hitting in the AL from the NL. To test this hypothesis, you allow the AL regression to have a different intercept and different slopes from the NL regression. You therefore create a binary variable for the American League (*DAL*) and estimate the following specification:

$$\widehat{Winpct} = -0.29 + 0.10 \times DAL - 0.100 \times teamera + 0.008 \times (DAL \times teamera) + 1.622 * ops - 0.187 * (DAL \times ops)$$
$$(0.12)\ (0.24)\qquad (0.008)\qquad (0.018)\qquad (0.163)\qquad (0.160)$$

With $R^2 = 0.92$, $SER = 0.02$.

What are the resulting regression equations for winning percentage in the AL and NL?

Next, calculate the t-statistics and say something about the statistical significance of the AL variables. Since you have allowed all slopes and the intercept to vary between the two leagues, what would the results imply if all coefficients involving DAL were statistically significant?

Question 6. Suppose you have the following result when you run a scatterplot of your dependent variable with one of your independent variable:

a) Give two suggestions of how you might modify your OLS regression equation from the base of $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

b) For both of the suggestions above, explain how to interpret the effect on Y of a change in X_1.