Ph.D. Econometrics I
Heinz School, Carnegie Mellon University
90-906, Spring 2000

Final

Instructions You may use any books, notes, calculators, and other aids you like. You may not converse, nor may you cooperate.

Please complete three of the four questions. Each question will count as 1/3 of the grade.

Please show all relevant work.

Please interpret your results in plain English.

1. See the attached description and codebook for the gallup dataset. See also the attached SAS program and output. Notice, `yrschool` is a proxy for years of schooling.

   Consider the following regression model (which we start with but are willing to consider modifying if necessary):

$$Salary = \beta_1 + \beta_2 yrschool + \beta_3 age + \epsilon \qquad (1)$$

   In answering the following questions, please be clear about exactly what model you are using, what page of SAS output you are using and why.

   (a) We are concerned about non-linearity in equation 1. Test for it.

   (b) Do salary paths *diverge* for people with different levels of schooling as they age?

   (c) Calculate the effect of a 1 year increase in age on salary, holding constant years of schooling, evaluating at sample means, if necessary. Also, provide a 95% confidence interval.

   (d) At what age is salary maximized (again evaluating at sample means if necessary). Provide a 95% confidence interval.

2. Consider the following dataset:

$$Y = \begin{bmatrix} -3 \\ -1 \\ -2 \\ -1 \\ 1 \\ 0 \\ 5 \end{bmatrix} \quad X = \begin{bmatrix} -3 \\ -2 \\ -1 \\ -0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

We wish to analyze the data using the following model:

$$Y = \beta_1 X + \epsilon \tag{2}$$

In answering the following questions, please feel free to use rounding in your calculations if that is convenient for you.

(a) Under the usual OLS assumptions, test the theory that X has no effect on Y.

(b) Test for heteroskedasticity.
Hint: the R^2 of a regression is not affected by using "de-meaned" RHS variables (i.e. using $X - \overline{X}$ instead of X), and this often makes inverting matrixes easier.

(c) Under the assumption of heteroskedasticity, test the theory that X does not affect Y.

3. Consider the (true) regression model:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \tag{3}$$

We can not see X_3, so the best we can do is to estimate as if the true model is:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon \tag{4}$$

In each situation below, tell me the following things (and explain):

- Does running OLS on equation 4 instead of on equation 3 cause bias in $\hat{\beta}_{1OLS}$ and $\hat{\beta}_{2OLS}$?
- If so, can we determine the direction of the bias?
- If so, what is the direction of the bias?

Throughout, assume $\overline{Y} = \overline{X_k} = 0$ — that all variables have sample mean 0 (this could be accomplished by "de-meaning" the data, for example). In addition, assume that $\beta_3 > 0$.

(a) If we know that $Cov(X_1, X_3) > 0$.

(b) If we know that $Cov(X_1, X_3) > 0$ and $Cov(X_2, X_3) > 0$.

(c) If we know that $Cov(X_1, X_3) > 0$, $Cov(X_2, X_3) > 0$, and $Cov(X_2, X_3) = 0$.

(d) If we know that $Cov(X_1, X_3) > 0$, $Cov(X_2, X_3) > 0$, and $Cov(X_2, X_3) < 0$.

(e) If

$$\hat{V}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 10 & -1 \\ 2 & -1 & 6 \end{bmatrix}$$

4. We are interested in measuring the average growth rate in Y over time. The coefficient β_2 in the following model is a useful approximation to the growth in Y over time:

$$\ln Y = \beta_1 + \beta_2 time + \epsilon \tag{5}$$

Assume that the error, ϵ is well-behaved.

One strategy for estimating β_2 is to run OLS on equation 5. Another strategy is to calculate the following variable:

$$\Delta \ln Y_t = \ln Y_t - \ln Y_{t-1}$$

and to take its mean.

(a) Is $\overline{\Delta \ln Y}$ a good estimator of the average growth rate of Y? Why or why not?

(b) Is the variance of $\overline{\Delta \ln Y}$ well estimated by the usual estimator of the sample variance of a mean, $\frac{1}{T-1} \sum \left(\Delta \ln Y - \overline{\Delta \ln Y}\right)^2$? Why or why not?

(c) Can you think of another, better estimator of the variance?

(d) Would OLS on equation 5 be a better estimator than $\overline{\Delta \ln Y}$? Why or why not?

General hint: The course was about regression analysis; therefore, this question must be about regression analysis.