Econ 551b
Problem Set 1

Prof. John Rust, Hiu Man Chan

Due: February 3, 1999

Question 1

a) Show that for any $k \times 1$ vectors X and a, $Var(a'X) = a'\Sigma a$ if $Var(X) = \Sigma$.

b) Prove that if X and Y are orthogonal matrices, so is XY.

c) Write the following quadratic form in matrix notation with a symmetric matrix and specify whether it is positive semi-definite:

$$-x^2 + y^2 + 3xy$$

Question 2

In this question, you are asked to study the effect of model misspecification through Monte Carlo Experiment.

a) Generate 500 samples, with each sample containing 1000 observerations from the nonlinear regression model:

$$y = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2) + \epsilon$$

where ϵ is distributed as standard normal, x_1 is distributed as $N(-1, 1)$, and x_2 is distributed as $N(1, 0.25)$. The true parameter values are set at $\beta_0 = 1$, $\beta_1 = 0.5$, and $\beta_2 = 1$.

b) For each sample generated, act as you don't know the true parameters and observe only x and y, estimate the parameters using OLS from the linear regression:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Therefore, you will obtain 500 vectors of OLS estimates from the 500 samples.

c) Compute summary statistics (such as mean, standard deviation, maximum, minimum) of your 500 OLS estimates for β_0, β_1, and β_2. Plot the empirical distribution of each of them. What can you say about the effect of model misspecification (estimating a nonlinear model as a linear one)?

Question 3

Extract data in file hrsdat.asc or hrsdat.dat (from the links on class web page), which contains 6851 observations from the Health and Retirement Survey. The data contains nine variables, as described in the file codebook.txt (also from links on class web page). You have to analyze the effect of education and other possible variables on earning power by estimating the linear regression model:

$$y = x\beta + \epsilon$$

where y is a measure of earning power, and x are factors that can affect earning power. You should:

a) State any assumption(s) you would like to make on ϵ in the regression model. Explain why you want to make the assumption(s).

b) Describe the dependent and independent variables used. This can involve manipulations of the variables in the data set. Report what these manipulations are. (E.g., creating a dummy which equals 1 if the respondent is white, and 0 if he/she is non-white from the variable "RACE".)

c) Obtain OLS estimates for β and standard errors of the estimates. Provide goodness of fit (such as R^2).

d) Interpret your results.