1. For the simple linear model $y = x\beta + u$ derive $\hat{\beta}$, the OLS estimator. What is needed to establish unbiasedness? Use this to show the unbiasedness of this estimator. What is needed to establish consistency of this estimator? Use this to show consistency of this estimator.

2. For this problem, use Stata. Assume that $\alpha = 1$, $\beta = 1$, $T=50$, and

$$y_i = \alpha + \beta x_i + u_i.$$  \hspace{1cm} (1)

Draw $u_i$ and $x_i$ as independent random variables each $\sim N(0,1)$ using a random number generator. Then generate your $y_i$ values in eq. (1). You have now generated $y$ and $X$ values that we can use for the following:

a. Obtain the OLS estimator of eq. (1) with this data. Using 399 bootstrap draws, write the Stata code to compute the wild and pairs bootstrap estimates of the standard error of $\hat{\beta}$ using the Program and Simulate commands (see my sample programs). Also run the canned Stata command to compute the bootstrap estimator. Refer to the bootstrap handout. Compare to the estimates obtained for the standard error of $\hat{\beta}$ computed using OLS.

b. Using the bootstrap estimates of $\hat{\beta}$ from part a, compute the bias-corrected OLS estimator. Does it differ much from the uncorrected OLS estimator? Why?

c. For equation (1), demonstrate that the OLS estimator, $\hat{\beta}$, is approximately unbiased using a Monte Carlo experiment with 1000 replications (remembering that 1000 should really be much larger). Assume that $X$ is fixed, that the true parameter values are as given above, and that $u$ is distributed as $N(0,1)$. Use $T=100$. Also, compare the sample variance of the estimates of $\beta$ to the average of the variances of $\hat{\beta}$ computed from the analytical formula. Redo all of this with $T=1000$. Are there any differences?

3. Consider

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i,$$ \hspace{1cm} (2)

where $\alpha = 0$, $\beta_1 = 2$, $\beta_2 = 2$, and $T = 100$.

a. Draw $x_{1i}$ and $x_{2i}$ so they have a bivariate normal distribution with correlation coefficient .35.

b. Draw $u_i$ as $N(0,1)$.

c. Generate $y_i$.

d. Now drop $x_{2i}$ from your regression model and compute the OLS estimator for

$$y_i = \alpha + \beta_1 x_{1i} + v_i,$$ \hspace{1cm} (3)
e. Show that the OLS estimator, \( \hat{\beta}_1 \), is biased using a Monte Carlo experiment with 1000 replications.

f. Now increase the sample size to \( T = 2000 \) and repeat this Monte Carlo experiment.

g. Is \( \hat{\beta}_1 \) consistent?

h. What necessary property for consistency has been violated?

i. What general name is given to the misspecified model in (5d)?

USE THIS LINK TO DOWNLOAD THE DATA NEEDED FOR THIS AND ALL OTHER PROJECTS: http://mitpress.mit.edu/9780262232586. Once at this page for the Wooldridge text, click on Supplemental Content in the left margin and then at the bottom on Download Stata files.

4. Using MROZ.RAW, reproduce the table 15.1 in Wooldridge.
   a. Use the canned Stata routines to estimate the Logit and Probit estimators. Then produce Logit and Probit estimates using hand-written ML routines. See my sample programs.
   b. Produce a separate table for each set of estimates.
   c. Now add a new interaction variable which you create by taking the product of Educ and Exper.
   d. After Probit and Logit, use the canned routine, Margins, and summarize the partial effects for each variable. Are there differences between models?

5. Using the same data as with the last problem, compute the OLS estimator.
   a. Show that the sample analogs of \( E(x'u) = 0 \) and \( E(u) = 0 \) hold true for the OLS estimator of the linear model, \( y = x\beta + u \).
   b. Compare estimated coefficients and standard errors with those obtained in the last question.
   c. What are the properties of this estimator?

GENERAL INSTRUCTIONS:

Do not hand in a printout with all the replications of the Monte Carlo and Bootstrap runs. Send me an electronic copy of your .do file and your key output where you have annotated it carefully. Clearly label each question as 2a, 2b, etc. using a large font. Highlight key text and make comments next to all relevant output using a text editor. Part of your grade will depend on how clearly and completely you do this. These same instructions apply to all projects in this course. Your best method to debug Stata is to print out lots of output—possibly after every new line of code. That is, print out everything. Comment out these commands only after you are sure that everything is correct. Every year this advisory is ignored by some students who waste large amounts of time over errors that can be easily detected by printing out everything until the entire program is
debugged. To debug your program, first run it with only 1 iteration—commenting out
the Program, Return, and Simulate commands—then add these back and do only say 5
iterations, then 100, and finally 1000. Initially use print statements literally after every
command. Comment these out and rerun the program before submitting your output to
me.