Computer Homework 7

This homework considers estimation of a linear regression model in the presence of strong correlation between the regressors. Use the data in the *Stata* data file cons.dta. This data file contains 46 quarterly observations of real personal consumption expenditure, c, and real disposable personal income, y, for the United States from 1990:Q1 through 2001:Q2. Also included in this file is a variable indicating the year and quarter, labeled time. Consider the following model:

$$\text{Model A:} \qquad c_t = \alpha + \beta y_t + \delta c_{t-1} + \epsilon_t$$

for t=1, .. ,T. It is not necessary to generate a new regressor corresponding to the lagged value of consumption. *Stata* will recognize **L.c** as the lagged value of c and will generate the necessary regressors internally. By default, *Stata* will delete any observations with missing values.

1. Use the *Stata* command **corr c y L.c** to get the correlation matrix for consumption, income, and lagged consumption. Does the correlation between regressors appear strong enough to cause a problem of near multicolinearity, based upon "rules of thumb" you have heard in the past?

2. Obtain OLS estimates of model A. Do these estimates appear to suffer from near multicolinearity?

3. It is sometimes suggested that near multicolinearity can be diagnosed by examining the sensitivity of OLS estimates to the elimination of small subsets of observations. Estimate the model for 1992:Q1 through 2001:Q2. This is done with the *Stata* command line

   reg c y L.c if tin(1992q1,2001q2)

   The **tin** function selects values of time between the specified limits. Repeat this for 1990:Q1 through 1999:Q2. How sensitive are the estimated coefficients to variation in the sample?