Professor R.L. Oaxaca | Assignment #5
Econ 418 | Fall 2009

Due in class Tuesday, December 1
(25 points)

The data for this assignment are contained in both the Excel files 'TRAFFIC2.xls' and 'MROZ.xls' and the STATA files 'TRAFFIC2.dta' and 'MROZ.dta' available at http://u.arizona.edu/~rlo. Be sure to attach the supporting computer print out to the completed assignment, show your work, and make clear where your answers are shown.

The following information pertains to questions 1 and 2 below. The TRAFFIC2 data set contains monthly time series data for the State of California over the period 1981 - 89. The variables of interest for this exercise are *totacc* (total number of statewide automobile accidents), *feb* - *dec* (dummy variables for each month), *spdlaw* (dummy variable =1 for each month after the 65 mph speed limit took effect), *beltlaw* (dummy variable = 1 for each month after the seatbelt law took effect), and $t$ (linear time trend for each month starting with 1 and ending in 108).

Some basic STATA commands that might be useful

> To inform STATA that time series data are being used and are ordered according to some linear time trend variable, say '*time*', type the command **tsset time**
>
> To create a one-period lagged value for some variable '*whoa*' and name the lagged variable '*whoaL*1', type the command **gen whoaL1 = whoa[_n-1]**
>
> In the time-series model $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$, to correct for first-order autcorrelation of the residuals using the two-step Cochrane-Orcutt method, type the command **prais Y X1 X2,corc twostep**

1. Consider the following model of traffic accidents

$$\begin{aligned} totacc_t &= \beta_0 + \beta_1\, feb_t + \beta_2\, mar_t + \beta_3\, apr_t + \beta_4 may_t + \beta_5 jun_t + \beta_6 jul_t + \beta_7 aug_t \\ &\quad + \beta_8 sep_t + \beta_9 oct_t + \beta_{10} nov_t + \beta_{11} dec_t + \beta_{12} spdlaw_t + \beta_{13} beltlaw_t + \beta_{14} t + u_t, \\ t &= 1, ..., 108 \end{aligned}$$

Assume that the error term is a normally distributed random variable with mean zero and constant variance and is independent of the regressors.

a. Suppose one suspects that $u_t = \rho u_{t-1} + \varepsilon_t$, where $\varepsilon_t$ satisfies all of the standard assumptions and $| \rho |< 1$. Use a consistent estimator to estimate $\rho$.

b. Use any appropriate test to test $H_0$: $\rho = 0$, $H_1$: $\rho \neq 0$ at the 5% level of significance. What can you conclude about the properties of *OLS* applied to the traffic accident model?

2. Use the Corchrane-Orcutt two-step $FGLS$ procedure to estimate the traffic accident model both manually and using the computer software automated procedure. Verify that the estimates are identical in the two cases.

   a. Do the results reveal any seasonal patterns to the number of traffic accidents? Explain. Interpret the estimated value of $\beta_{14}$

   b. Compare the $OLS$ results with the $FGLS$ results in terms of how inferences about the effects of the regressors might differ between the two sets of results.

The following information pertains to questions 3 below. The MROZ data set is a random sample of 753 married women interviewed in 1976. The variables of interest for this exercise are $hours$ (hours worked in 1975), $wage$ (hourly wage), $huswage$ (husband's hourly wage), $hushrs$ (hours worked by husband in 1975), $age$ (age in years), $kidslt6$ (# of kids < 6 years old), $kidsge6$ (# of kids 6 - 18 years old), $exper$ (labor market experience in years), and $educ$ (years of schooling).

Some basic STATA commands that might be useful

> To drop all observations for which some variable '*money*' equals 0, type the command **drop if money ==0**
>
> To use instrumental variables (IV)/Two Stage Least Squares (2SLS) to estimate the model $Y_{1t} = \gamma_{10} + \gamma_{11} X_{1t} + \beta_{12} Y_{2t} + u_{1t}$ where it is believed that $Y_{2t}$ is correlated with $u_{1t}$ and that the variables $X_{2t}$ and $X_{3t}$ are suitable identifying instruments for $Y_{2t}$, i.e., are correlated with $Y_{2t}$ but not correlated with $u_{1t}$, type the command **ivreg Y1 X1 (Y2 = X2 X3)**

3. Consider the following structural labor supply function for the 428 married women who worked in the market sector in 1975, i.e. those women for whom $hours > 0$:

$$\ell n(hours_i) = \gamma_{10} + \beta_{11} \ell n(wage_i) + \gamma_{11} \ell n(hearn_i) + \gamma_{12} age_i + \gamma_{13} kidslt6_i + \gamma_{14} kidsge6 + u_{1i}$$

where $hearn_i = huswage_i * hushrs_i$.

   a. Consider the possibility that $\ell n(wage_i)$ may be endogenous and correlated with $u_{1i}$.

      (1) Estimate the labor supply function by $IV/2SLS$ using $exper_i$ and $educ_i$ as identifying instruments for $\ell n(wage_i)$.

      (2) Show that the coefficient estimates you obtain by manually carrying out the $IV$ estimation are identical to those obtained from the computer software automated version. Explain why the estimated standard errors differ between the manual and the automated procedures.

   b. Now consider a second structural equation for wages given by

$$\ell n(wage_i) = \gamma_{20} + \beta_{21} \ell n(hours_i) + \gamma_{21} exper_i + \gamma_{22}\, educ_i + u_{2i}$$

   You now have a two-equation simultaneous equations model.

(1) Specify which variables of the two-equation model are endogenous and which variables are predetermined (exogenous).

(2) Use the order condition to determine the identification status of each of the two structural equations.

(3) Estimate the wage equation by $2SLS$. Based on the empirical results for the wage and hours equations, what would you conjecture about the properties of $OLS$ applied to these two structural equations?

.