Microeconomics

Assignment 3

1. Consider the treatment-outcome model

$$y = x\beta + \alpha d + \epsilon, \tag{1}$$

where d is a binary indicator variable taking the value 1 if treatment is assigned and 0 if treatment is not assigned.

(a) Is randomized treatment a sufficient condition for identification of α and β?

(b) If participation of treatment ($d = 1$ or 0) is correlated with y, suggest a method to obtain α.

(c) Suppose the outcome equation (eq. (1)) is unknown, but eligibility of treatment is exogenously determined, suggest a method to obtain the average treatment effect.

2. For this exercise you will use the data set WAGES.DTA taken from the Panel Study of Income dynamics, Cornwell and Rupert (*Journal of applied Econometrics*, 1988). The data set contains observations on 595 individuals over 7 years, 1976-1982.

The list of the variables is as follows:

ID= variable identifying individuals

YEAR= variable identifying time

EXP= Years of full-time work experience.

WKS= Weeks worked.

OCC= (OCC=1, if the individual is in a blue-collar occupation).

IND= (IND=1, if the individual works in a manufacturing industry).

SOUTH= (SOUTH=1, if the individual resides in the South).

SMSA= (SMSA=1, if the individual resides in a standard metropolitan statistical area).

MS= (MS=1, if the individual is married).

FEM= (FEM=1, if the individual is female).

UNION= (UNION=1, if the individual's wage is set by a union contract).

ED= Years of education.

BLK= (BLK=1, if the individual is black).

LWAGE= Logarithm of wage.

The goal of this exercise is to introduce you to the analysis of panel data sets. You will have the opportunity to implement different specifications (random effects, fixed effects, etc.), compare them and choose the most appropriate model. All STATA commands for panel data begin with the prefix "*xt*". Before you start answering the questions, it would be useful to look for "*xt*" command in the STATA help.

1) Explore the data set. Identify which variables are time variant, individual variant or both.

2) Define the data as a panel structure in STATA. Specify the variable that represents the individual and the variable that represents the year. Compute summary statistics for all the variables. Are there any 'suspicious values'?

3) Construct the square of work experience, EXP2. Regress by GLS the logarithm of wage rate on a constant term, EXP, EXP2, WKS, OCC, IND, SOUTH, SMSA, MS FEM, UNION, ED and BLK. Do the signs of the parameters agree with your intuition? What variables are significant at 5% level? Is this an appropriate procedure?

4) Implement the following methods for the wage equation:

   a) Fixed effects (within)

   b) Random effects (GLS)

   c) Between

   d) Random effects (ML)

For each one of these estimations answer the following questions:

Do the signs of the parameters agree with your intuition? What variables are significant at 5% level? Is the estimation procedure appropriate for the context? Is the estimator consistent? Compare the results. Do you accept or reject the presence of individual effects? Why or why not?

5) Compute the estimates of specific effects (fixed and random). Comment.

6) Would you prefer the fixed effects or the random effects results? Why? (you can use the Hausman test).

3. Consider the model

$$y_{it} = \mu + \bar{x}_i a + x_{it}\beta + \omega_i + u_{it}, i = 1, \ldots, N,$$
$$t = 1, \ldots, T,$$

where $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, Cov $(\omega_i x_{it}) =$ Cov $(u_{it}, x_{it}) = 0, \omega_i$ is independently distributed over i with mean 0 and variance σ_α^2, and u_{it} is independently, identically distributed over i and y with mean 0 and variance σ_u^2.

(a) Use the formula of partitioned inverse to show that the generalized least squares (GLS) estimator of (μ, β, a) as

$$\hat{\mu} = \bar{y} - \bar{x}\hat{\beta}_b,$$
$$\hat{\beta} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - x_i)^2\right]^{-1}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - y_i)\right]$$
$$\hat{a} = \hat{\beta}_b - \hat{\beta},$$

where

$$\hat{\beta}_b = \left[\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2\right]^{-1}\left[\sum_{i=1}^{N}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})\right],$$
$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N}\bar{x}_i, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N}\bar{y}_i, \bar{y}_i = \frac{1}{T}\sum_{t=1}^{T}y_{it}.$$

(b) When $a = 0$, show that the GLS estimator of (μ, β) is equivalent to the least squares regression of $[y_{it} - (1 - \psi^{\frac{1}{2}})\bar{y}_i]$ on a constant and $[x_{it} - (1 - \psi^{\frac{1}{2}})\bar{x}_i]$ and the Var $(\hat{\beta}) = \sigma_u^2\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)^2 + T\psi\sum_{i=1}^{N}(\bar{x}_i - \bar{x})^2\right]^{-1}$, where $\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$.