Problem Set #8: Panel Data (updated 11/2/2011)

Economics 435: Quantitative Methods

Fall 2011

Please use WebCT to turn in both the Word document and R code for question (1) and turn in your answer to question (2) in class.

1 Public sector unionization and the size of government: Part V

a) Fill in Table 4. Each regression you estimate for Table 4 will have the following feature:

- It will be based on a panel of states for the years 2000 through 2008.
- It will include state-level fixed effects. These will not be reported in the table.
- It will include year fixed effects. These will not be reported in the table.
- The only explanatory variable (other than the fixed effects) will be the public sector union membership rate.

Here's a work plan for estimating the fixed effects model in R:

1. First, adjust your existing code so that it can be used to easily read in and process data from other years. My suggestion is:
   (a) Separate the part of the code that creates your data set from the part that estimates regressions or makes figures.
   (b) Put the data-creation code into a function (use R's `function()` command) that takes the current year as an argument, and outputs a data set for that year.
   (c) Here's some code that will help you. Suppose you the variable `year` contains the year, and that you want to create a string by pasting together the word "State U_" and whatever number is stored in `year`. Then you can do this with the function `paste`: the exact code is `paste("State U_",year,sep="")`
   (d) Add a variable to the data set identifying the current year.
   (e) Test the function by using a call to the function to read in your three existing data files and estimate the regressions.
2. Next, create the CSV files needed for all the years. In order to facilitate them being read in by the function you just created:
   (a) The format,including the column headings, should be exactly identical.

(b) The filenames should be systematic (for example the unionization file for year XXXX should be called (`State U_XXXX.csv`).

(c) Test to make sure your new CSV files are in the right format by running your data creation and estimation code on them.

Since this task is labour-intensive and doesn't teach you much, I've done it for you. The CSV files are available in the ZIP file `unionization.zip`.

3. Using data-processing function you have created, read in all of the years of data and combine them into a single data set using the `rbind()` function.

4. R does not have built-in commands for panel data. Instead, those commands are in an external "package" called `plm` (for panel linear models). A package is simply a set of functions and datasets that are used for a common set of tasks. We need to load the `plm` package.

   (a) If you are using a campus computer, the `plm` package is already installed. It can be "loaded" by executing the command `library("plm")`.

   (b) If you are using your own computer the `plm` package may or may not be installed. If it isn't installed, you will get an error message when you try to execute `library("plm")`. In that case, you can install it by executing the command `install.packages("plm")`. Then you can load it by calling `library("plm")`.

5. Next,estimate your regressions using the command `plm()`. Be sure to set the `effect=` option to `effect="twoways"`, and be sure to use the `index=` option correctly.

b) Update the discussion in Section 3 to include these regressions.

c) Fill in Section 4.3.

2 Fixed effects with measurement error

Consider the following model of fixed effects with measurement error.

Suppose that we have a panel data set consisting of a random sample of individuals (indexed by $i \in \{1, 2, \ldots, n\}$) observed at two points in time (indexed by $t \in \{0, 1\}$. We observe the outcome y_{it}, which is a function of an individual-specific fixed effect a_i, a time-specific fixed effect δ_t, a single time-varying explanatory variable x_{it}, and a time-varying unobserved component u_{it}:

$$y_{it} = a_i + \delta_t + \beta x_{it} + u_{it} \tag{1}$$

where u_{it} is mean-zero, IID across time and individuals, and independent of all other exogenous variables.

To keep things simple, we suppose that:

$$x_{it} = x_i + v_{it} \tag{2}$$

where v_{it} is mean-zero, IID across time and individuals, and independent of all other exogenous variables.

However, we do not get to observe x_{it}, we only get to observe a noisy measure of it called $\tilde{x}_{it}$. Suppose that:

$$\tilde{x}_{it} = x_{it} + \epsilon_{it}$$

where ϵ_{it} is mean-zero, IID across time and individuals, and independent of all other exogenous variables.

Let:

$$\begin{aligned}
\mu_x &= E(x_i) = E(x_{it}) \\
\mu_a &= E(a_i) \\
\sigma_a^2 &= var(a_i) \\
\sigma_x^2 &= var(x_i) \\
\sigma_v^2 &= var(v_{it}) \\
\sigma_\epsilon^2 &= var(\epsilon_{it}) \\
\rho_{a,x} &= corr(a_i, x_i)
\end{aligned}$$

We are interested in estimating β, the effect of x_{it} on y_{it}. However, there are two problems with doing so: correlation between x_{it} and the individual-specific fixed effect (a_i), and measurement error in $\tilde{x}_{it}$.

Let the OLS estimator for β be defined as

$$\hat{\beta}^{OLS} \equiv \frac{c\hat{o}v(y_{it}, \tilde{x}_{it})}{v\hat{a}r(\tilde{x}_{it})}$$

and let $\beta^{OLS} = \text{plim } \hat{\beta}^{OLS}$. Let the FD (first difference) estimator be defined as:

$$\hat{\beta}^{FD} \equiv \frac{c\hat{o}v(\Delta y_i, \Delta\tilde{x}_i)}{v\hat{a}r(\Delta\tilde{x}_i)}$$

where $\Delta\tilde{x}_i = \tilde{x}_{i1} - \tilde{x}_{i0}$ and $\Delta y_i = y_{i1} - y_{i0}$. Also let $\beta^{FD} = \text{plim } \hat{\beta}^{FD}$.

a) Find the probability limit of $\hat{\beta}^{OLS}$ in terms of the parameters:

$$(\mu_x, \mu_a, \beta, \sigma_a^2, \sigma_x^2, \sigma_u^2, \sigma_v^2, \sigma_\epsilon^2, \rho_{a,x})$$

b) Find the probability limit of $\hat{\beta}^{FD}$ in terms of these same parameters.

c) Suppose (for this part of the question only) that there is no fixed effect, i.e., $\sigma_a^2 = 0$. Find each of the two probability limits. Which estimator produces a smaller (or at least no larger) asymptotic bias?

d) Suppose (for this part of the question only) that there is a fixed effect, but that it is uncorrelated with x_i, i.e., $\rho_{a,x} = 0$. Find each of the two probability limits. Which estimator produces a smaller (or at least no larger) asymptotic bias?

e) Suppose (for this part of the question only) that there is no measurement error, i.e., $var(\epsilon_{it}) = 0$. Find each of the two probability limits. Which estimator produces a smaller (or at least no larger) asymptotic bias?

f) Consider the following two statements (difference between the two is in **boldface**):

> When estimating a model using panel data in which the explanatory variable is measured with error, a fixed effects estimator will be preferable (in terms of having smaller asymptotic bias) to a simple OLS estimator when the covariance between the fixed effect and the explanatory variables is **large** relative to the measurement error.
>
> When estimating a model using panel data in which the explanatory variable is measured with error, a fixed effects estimator will be preferable (in terms of having smaller asymptotic bias) to a simple OLS estimator when the covariance between the fixed effect and the explanatory variables is **small** relative to the measurement error.

Based on your results, which of the two statements is correct?