Problem Set 5

This problem set will revise some of the material covered in Handout 5 on panel data models. This will require you to familiarize yourself with Stata's panel-data commands.

```
help xtset
help xttab
help xtreg
```

You will be using a dataset that comes with Stata: psidextract.dta. The data is a correct version of the PSID sample in Cornwell and Rupert (1988), found in Baltagi and Khanti-Akom (1990). It includes a sample of 595 individuals observed for the years 1976-82.

Preamble

```
<IPython.core.display.HTML object>
```

Create a do-file for this problem set and include a preamble that sets the directory and opens the data. For example,

```
clear
//or, to remove all stored values (including macros, matrices, scalars, etc.)
*clear all

* Replace $rootdir with the relevant path to on your local harddrive.
cd "$rootdir/problem-sets/ps-5"

cap log close
log using problem-set-5-log.txt, replace

use problem-set-5-data.dta, clear
```

Questions

1. Set the unit identifier and time variable using `xtset`. Note, you can also use `tsset` for this task. This will allow you to use `xt` package commands.

2. Describe and summarise the variables in the dataset using the normal `describe` and `summarize` commands.

3. Describe and summarise the variables in the dataset using the panel commands: `xtdescribe` and `xtsummarize`. Comment on the information provided.

4. Use the command `xttab` and `xtrans, freq` to describe transitions over time in the variable `south`.

5. Create the variable: `expsq=exper^2/1000`. Why would you scale the variable in this way?

6. Estimate the following model using pooled OLS, between-group, feasible GLS, within-group, LSDV, and first-difference. For the first-difference estimator, you can define a first-difference in Stata using the time-series operator: `D.variable`.

$$\ln(Wage_{it}) = \beta_1 + \beta_2 Exper_{it} + \beta_3 Exper_{it}^2 + \beta_4 Weeks_{it} + \beta_5 Eduyrs_{it} + \varepsilon_{it}$$

With each model, store the results using `estimates store`. For example,

```
* clear existing stored estimates
est clear

* Pooled OLS
regress lwage exper expsq weeks ed
est store OlS

* alternatively,
eststo OLS: regress lwage exper expsq weeks ed
```

7. Using the formula from Handout 5, replicate the value of θ reported above by the FGLS estimator. Note, you will need to use the stored values of σ_ε^2 and σ_α^2.

8. Make a table of the computed estimates. You can either use `estimates table` or `esttab`. The latter is part of the `estout` package, which you may need to install: `ssc install estout`.

9. Perform a Hausman test comparing the results of the FLGS and WG estimators. You should use the `hausman` command, with the option `sigmamore`. Be sure to get the order of the estimates correct. What do you learn from the test?

10. Estimate FGLS for the model below:

$$\begin{aligned}\ln(Wage_{it}) =& \beta_1 + \beta_2 Exper_{it} + \beta_3 Exper_{it}^2 + \beta_4 Weeks_{it} + \beta_5 Eduyrs_{it} \\ &+ \gamma_2 \overline{Exper}_i + \gamma_3 \overline{Exper^2}_i + \gamma_4 \overline{Weeks}_i + \varepsilon_{it}\end{aligned}$$

You will need to manually create the variables: $\{\overline{Exper}_i, \overline{Exper^2}_i, \overline{Weeks}_i\}$ - the individual-level averages of each variable. This is referred to as the Mundlack correction. Once you have

estimated the model, repeat the Hausman test comparing these results with those of the WG estimator. What is the significance of the Mundlack correction?

11. Export the results as a single CSV/Excel file. You can use `esttab` for .csv or `outreg2` for .xlsx.

Postamble

```
log close
```