

## Problem set 15: week 24

This problem set is to be completed prior to your Stata class. The data set used `sleep.dta` is available on the EC203 website, please copy it on to a memory stick or your H: drive. I will highlight the main commands, use either: **help**, **findit** or the **search** function for guidance on each.

1. Open a do file in Stata. All the commands we use in this problem set will be copied into here. This is so you can recall what we have done, and the analysis can be repeated. It will also be useful for you to annotate the do file as you go along.
2. Load `sleep.dta` into Stata. We are interested in estimating the relationship between hours slept and hours worked per week. Given the variables in the data set, write out a panel data model for total number of minutes slept per week. Call this model A.
3. Use pooled OLS (POLS) to estimate your model. I.e. just use the standard regression command **reg slpnar 'controls'**. Interpret the coefficient of total minutes worked per week.
4. With reference to individual unobserved heterogeneity explain why we may be concerned about bias in the above regression.
5. We consider two main ways to deal with unobserved heterogeneity: first differences and fixed effects. Estimate the model using first-differences. Estimate the model using fixed effects (this requires you to use **xtreg slpnar 'controls', fe**).
6. When we had cross-sectional data we typically assumed there was zero correlation between the error terms. Why was this a reasonable assumption in the cross-sectional case? Explain why for panel data zero correlation in the errors is a highly dubious assumption to make.
7. One way to deal with non-zero correlation in the errors is to allow arbitrary correlation within clusters (individuals in this case). Using the **cluster** option on the regression command estimate model A clustering by individual. Compare the coefficient and standard errors between the POLS and the clustered POLS.
8. An alternative way to deal with non-zero correlation in the errors is to use random effects. Remembering to **xtset id year** your data,<sup>1</sup> re-estimate model A using random-effects

---

<sup>1</sup>This informs Stata you are using panel data.

(this requires you to use **xtreg slpnar 'controls', re**). Compare the coefficient and standard errors between the POLS, clustered POLS and random effects.

**9.** Using all the information above, which model do you think has the most reliable estimates of the causal effect of work on sleep.

**10.** Annotate and close your do file.