**Problem 3 [35 points]** In this this question you will calculate and review a fixed effects model and a fixed effects model with time effects. The context and data for the question come from work that one of your instructors did as a statistical consultant to the United Nations in 2015.

One of the major challenges that the poorest of the poor countries in the world face in working to increase living standards is that couples do not have access to birth control, so they cannot control the initial timing, the spacing, and the quantity of children. The list of problems that results from this situation is long and often obvious, involving lack of per capita resources to education, health care and infrastructure. It also involves the physical and social damage of child marriage and early birthing, of the early aging and death of overtaxed mothers, and the abandonment of women and children.

The United Nations, as well as other world entities (think Melissa Gates) works with donor countries to put contraceptives within reach of couples in poor countries. Often these couples do not even know that such a thing as contraception is possible. The UN subdivision that handles this work is called the UN Population Fund, and goes by its initials in French, UNFPA.

Donor countries want to know if the money going to UNFPA is having an effect.

Data about life in poor countries is hard to come by, but usually there is an internationally-assisted small survey in these countries every handful of years. UN demography professionals turn these surveys into annual estimates via a somewhat involved Bayesian technique. Thus we have annual data, albeit somewhat second-hand.

Contraception availability is measured via the contraceptive prevalence rate, CPR, sometimes written as mCPR because some countries do not want contraception involving unmarried women to be discussed. The goal is to increase the CPR in poor countries. Additionally, there is an education process to teach couples that contraception is possible and in a perfect world they may even have a choice of methods. This education process creates “demand” for contraception, which can then be characterized as met or unmet. So in addition to raising the CPR, the UNFPA wants to raise total demand.

CPR and total demand increase slightly over time even in very poor countries as some contraception resources are imported by a country’s elite and perhaps middle class. The idea of international involvement is to increase the *rate of change* CPR and total demand increases.

The UNFPA must be invited into a country by its government. Some governments instead invite in other international entities, or otherwise align away from UNFPA, this providing a non-UNFPA-assisted control group that can be referenced. Before considering this control group, your instructor looked at the rate of change in CPR and total demand in countries in which the UNFPA was active, 12 in total, on the basis of comparing the five years before involvement and the five years aver involvement. That included a comparison of means (SW chapter 3) and a fixed effects model, which we do not replicate in this problem set for brevity. That type of analysis is not well-suited to using time effects because *all* the countries get the treatment at year six, so the time effect in year six gets mixed with the UNFPA effect. Even if the actual year of first UNFPA involvement is used instead of a uniform five years before and five years after, most of the counties started UNFPA involvement in the same year, so the dispersion is inadequate for the use of time effects.

The control group is 57 other poor countries with another, presumably less effective, alignment. In total the dataset is 69 countries over the 10 years of 2004 to 2013. A draft working paper from 2015 is included for any student interested. The analysis replicated in this problems set begins on page 23.

Recall that the software is case sensitive. The data set is composed of Country, not country, Year, not year, CPR, TD, and UNFPA = 1 if the UNFPA is involved. So there is only one regressor; there are no control variables. The one regressor is binary. The data are arranged in the conventional panel data set way. The first entity is listed for its first time period, then its second … to its last, then the second entity is listed for its first time period, etc. Note that CPR and total demand are measured in percentages in the format of 50 represents 50%, so a number like 0.33 means increasing the rate of change of change of the dependent variable by 1/3 of 1% faster than it would otherwise be.

The model we will use is:

where,

total demand (or CPR), indexed by country and year

coefficient of interest, measuring the UNFPA Supplies effect, listed as UNFPA in the results

an indicator variable, 0 if UNFPA Supplies is not involved, 1 if UNFPA is involved

the estimated fixed effect for country i

random disturbance

You will be asked to reproduce the panel results in the table below, augmenting the table with information that was not part of it when it was originally created. You will be asked to write (type) in the standard errors below the estimated coefficients, and indicate the significance level of the coefficients, if any, by placing ‘\*’ to the right of the estimate if it is significant at the 10% level, ‘\*\*’ if it is significant at the 5% level, and ‘\*\*\*’ if it is significant at the 1% level. Additionally, your software code and responses, as well as your written answers, are requested to be written below each question part as usual.

Table 3. The average increase in the rate of change of total demand and CPR from UNFPA Supplies involvement is about one quarter and one third of a percent respectively.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Comparison | Fixed | Fixed Effects with |  |  |  |
|  | of Means | Effects | Time Effects | Average |  |  |
|  |  |  |  |  |  |  |
| CPR | 0.265 | 0.22 | 0.253 | 0.246 |  |  |
|  |  | **\_\_\_\_\_\_** | **\_\_\_\_\_\_\_** |  |  |  |
| Total Demand | 0.299 | 0.326 | 0.376 | 0.334 |  |  |

**\_\_\_\_\_\_** **\_\_\_\_\_\_\_**

Some preliminary thoughts: In R, you may choose to use the plm package. Occasionally some R users have trouble installing the plm package. If you experience difficulty, you might try using the following line of code before (re-)attempting the install. options("install.lock"=FALSE)

In both software packages, you must be conscious that the software will eventually need to know that Country is the first index and Year is the second index, either as a setup command or within the regression call. Always used clustered standard errors, by Country (entity) of course. The usual fixed effects approach (involving estimation by de-meaning within an entity) is sometimes called a “within” model. The rarely-used alternative is a “between” model.

(a) [5 points] Run the fixed effects regression for Total Demand. Enter the standard error above, mark the table with asterisks indicating significance, if appropriate, and paste your software work below. A Stata note: Rather than not reporting a constant, or reporting a constant of zero, Stata adjusts the fixed effects values (which stay in the background, unreported) by subtracting them from their mean, and reporting the mean as the constant.

(b) [5 points] As a point of comparison, run the fixed effects model in part a without the portion of the command that requests the clustered standard error. Paste your results below. Is the p-value of the estimated coefficient on UNFPA noticeably lower, and the t-statistic higher?

(c) [5 points] Now add time effects to the model in part a. Time effects are often expected to be less notable than entity effects, so in some software the command process to include them is not as streamlined. You can always revert to adding dummy year variables to the regression if desired, leaving out one of your choosing. The model is depicted below.

where the variables are as before except that,

the estimated effects of each year on the countries in both the UNFPA and control groups.

Enter the standard error above, mark the table with asterisks indicating significance, if appropriate, and paste your software work below. Comment qualitatively on whether the level of the coefficient of interest or its p-value have seemed to change much, in your opinion.

(d) [5 points] Test the time effects (time dummies) for significance jointly. What is p-value and the decision regarding the null hypothesis at the 5% significance level? Include your software work below.

(e) [5 points] Now run the a fixed effects model, and then a fixed effects model with time effects, using CPR as the dependent variable. Enter the standard errors above, mark the table with asterisks indicating significance, if appropriate, and paste your software work below. Comment qualitatively on whether the level of the coefficient of interest or its p-value have seemed to change much between the two regressions, in your opinion.

(f) [5 points] Again, test the time effects (time dummies) for significance jointly. What is the p-value and the decision regarding the null hypothesis at the 5% significance level? Include your software work below.

(g) [5 points]Suggest one or more control variables that do not fall into either of the two categories below.

- different across entities but does not vary over time

- varies over time but is the same for all entities