This homework will familiarize you with executing and interpreting difference-in-difference (DiD) designs, and basic applications of fixed effects. The first question pertains to DiD, the second to fixed effects.

The DiD pertains to a policy intervention in India which nudged banks to open branches in financially under-developed areas. We shall explore whether the policy affected firm capital. The policy was introduced in 2005 and defined some districts as "underbanked". These will be our "treated" districts, with the treatment starting in 2005. Districts in India are equivalent to U.S. counties. No district was "treated" prior to 2005.

Some shameless self-promotion: you can learn more about the policy in this paper – `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4349526`

Note: use the sample weights for all summary statistics and regressions. The weights are defined by the variable mult.

Note: use hw7d1 for this question.

Q.1a) We will begin with an "unconditional" DiD, where we only compute the difference in sample means between the treated and control groups. The data has two years, 2004 and 2007. Calculate the mean value of logged capital stock ($\ln\_avg\_nfa$) for four groups: (i) treatmeant, before policy; (ii) treatment, after policy; (iii) control, before policy; (iv) control, after policy. Use these sample means to calculate the "unconditional" DiD coefficient.

Note: treatment status is defined by the binary *underbanked* variable; the post-treatment period is defined by the binary *post_treat* variable.

Q.1b) Conisder the following DiD regression:

$$\ln Capital_{idt} = \beta_0 + \beta_1 Underbanked_d + \beta_2 Post_t + \beta_3 Underbanked_d \times Post_t + \epsilon_{idt}$$

The unit of observation is firm i, located in district d, in year t. What information is provided

by the β_0 and β_1 coefficients. What is the "treatment" effect of the policy intervention on capital investment.

Q.1c) List two factors which might possibly be correlated with a district being "under-banked" (financially under-developed with less banks prior to 2005) and firm capital. How do you think these might affect the treatment effect?

Q.1d) Now we shall "flexibly" control for select factors. Start by including state fixed effects. Mention 2 factors which might be accounted for by state fixed effects, and 2 factors which will not be accounted for by state fixed effects. How does β_1 and β_3 change once you include state fixed effects? State identifiers are in variable *stcode*.

Hint: consider using the areg command for this.

Q.1e) Replace the state fixed effects with district fixed effects. How does β_3 change once you include district fixed effects? Why is β_1 no longer reported with district fixed effects?

Q.1f) Replace the district fixed effects with firm fixed effects. Mention 2 factors which might be accounted for by firm fixed effects. What type of factors are not accounted for by firm fixed effects? How does β_3 change once you include firm fixed effects? Firm identifiers are in variable *ID*.

Q.1g) After including firm fixed effects, what is the assumption needed for the treatment effect (β_3) to be unbiased?

Q.1h) As financially under-developed areas were eligible for treatment, household consumption and underbanked status is likely to be correlated. Using the *corr* command in *Stata*, check if the *underbanked* dummy is positively or negatively correlated with households' monthly consumption (*mpce30*).

Q.1i) Control for the interaction of household consumption (measured in 2004) with the post-treatment indicator (*mpce30_post*). How does the estimate of β_3 change if you include this interaction term? Based on your results, is there any alternate explanation of the treatment effect, aside from the entry of banks in underbanked areas? Include firm fixed effects in this estimation.

Q.1j) Another variable which is likely to be correlated with underbanked status is the share of rural households in districts. Include the share of rural households, interacted with the post-treatment indicator (*rural_hh_post*). How does β_3 change if you include this additional control variable? What inference can you draw about the treatment effect from this regression? Include firm fixed effects in this estimation but not *mpce30_post*.

Q.1k) Redo the exercise in Q.1f) but now use the *robust* command in *Stata* to have robust standard errors. How does the treatment effect change? Compare both the estimated $\hat{\beta}_3$ coefficient, and $se(\hat{\beta}_3)$ with the estimates in Q.1f).

Q.1l) Redo the exercise in Q.1f) but now cluster the errors by district. How does the treatment effect change? Compare both the estimated $\hat{\beta}_3$ coefficient, and $se(\hat{\beta}_3)$ with the estimates in Q.1k). The district identifier is *stdist*.

Q.2) This exercise will explore the relationship between bank branches and firm profits over time, after accounting for time-invariant differences across geographic units and firms. Use the dataset *hw7d2* for this exercise.

a) Consider the regression specification

$$ShProfit_{idt} = \beta_0 + \beta_1 BranchPC_{dt} + \epsilon_{idt}$$

Estimate this using OLS and interpret the β_1 coefficient. What is the assumption for the β_1 coefficient to be unbiased?

b) Consider year fixed effects. List two factors which are being accounted for by year fixed effects. Include year fixed effects using the command *i.year* in *Stata*. How does β_1 change upon including year fixed effects. Continue to include year fixed effects in all specifications from now.

c) Districts with more bank branches are possibly more politically conscious as they demand more financial infrastructure. How does the omission of political consciousness bias the estimated β_1?

Hint: your answer would depend on your assumption about how political consciousness in a district affects firm profits.

d) Re-estimate a) after including district and year fixed effects which can potentially account for political consciousness in an area. How does the β_1 coefficient change? When would the district fixed effects not fully account for local political consciousness? The district identifier is *stdist*.

Hint: use the areg command for this and absorb the district identifiers

e) Re-estimate a) after including firm and year fixed effects and report the β_1 coefficient. What is the assumption needed for β_1 to be unbiased, after including firm fixed effects. Which state/district characteristics are accounted for by firm fixed effects, and which are not? List one firm-level factor which might not be accounted for by firm fixed effects and bias the estimated β_1 coefficient. The firm identifier is *ID*.

f) Now add in 4 time-varying controls: firm age (*age*) and its squared (*sq_age*), whether the firm is listed (*listed*), whether the firm imports inputs (*importer*), and total number of workers hired by the firm (*nototalworker*). How does the β_1 change upon including these controls? Continue to include firm and year fixed effects and weight your regression using multiplier weights (*mult*).

g) Cluster your standard errors by district – the level at which the independent variable of interest varies using *Stata*'s cluster command. Do bank branches have a statistically significant impact on firm profits at the 1% level after clustering of standard errors?