Tutorial 4

Fixed Effects

Matriculation numbers: 3391610

3290592

1. REGRESSION SET 1

a) Run a regression of logGDP_future on private_credit_past with country fixed effects using the Stata command reghdfe. For now, don't pay attention to the standard errors. Comment briefly on the sign and size of the coefficient. From now on, keep only observations included in that first regression.

```
HDFE Linear regression                            Number of obs   =        695
Absorbing 1 HDFE group                            F(   1,    664) =      63.20
                                                  Prob > F        =     0.0000
                                                  R-squared       =     0.3869
                                                  Adj R-squared   =     0.3592
                                                  Within R-sq.    =     0.0869
                                                  Root MSE        =     5.2544
```

logGDP_fut~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
private_c~st	-.1186336	.0149227	-7.95	0.000	-.1479349	-.0893322
_cons	9.311867	.2298812	40.51	0.000	8.860486	9.763249

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
CountryCode	30	0	30

The regression results suggest that there is a negative relationship between private credit and GDP. On average, private credit decreases lending by 1.8% of the standard deviation of log of GDP.

$\beta/\sigma(y) = -0.018$

b) Run the same regression as in a) using the Stata command reg and including one dummy variable for each country. Compare the coefficient obtained and the R 2 measures to the previous regression

Firstly, we dropped missing values from the data. This time we used another method of including fixed effects, namely including dummy variables for each country. As shown in the below, we get the exactly the same coefficient and R-squared.

Source	SS	df	MS	Number of obs	=	695
				F(30, 664)	=	13.97
Model	11569.7946	30	385.659819	Prob > F	=	0.0000
Residual	18332.254	664	27.6088163	R-squared	=	0.3869
				Adj R-squared	=	0.3592
Total	29902.0486	694	43.0865253	Root MSE	=	5.2544

logGDP_fut~e	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
private_c~st	-.1186336	.0149227	-7.95	0.000	-.1479349	-.0893322
CountryCode						
40	-5.358042	1.861452	-2.88	0.004	-9.013084	-1.703
56	-3.859521	1.419179	-2.72	0.007	-6.64614	-1.072901
124	-2.538319	1.306184	-1.94	0.052	-5.103067	.0264289
203	-2.259734	1.875434	-1.20	0.229	-5.942229	1.422762
208	-5.794383	1.814033	-3.19	0.001	-9.356315	-2.232452
246	-3.497618	1.315223	-2.66	0.008	-6.080115	-.9151199
251	-5.179063	1.383824	-3.74	0.000	-7.896261	-2.461864
276	-5.139627	1.318308	-3.90	0.000	-7.728182	-2.551072
300	-4.906199	1.805055	-2.72	0.007	-8.450503	-1.361895
344	-.0315171	1.63953	-0.02	0.985	-3.250805	3.187771
348	-3.83051	1.606197	-2.38	0.017	-6.984346	-.6766736
360	5.891957	2.550359	2.31	0.021	.8842167	10.8997
372	-7.251999	2.945056	-2.46	0.014	-13.03474	-1.469255
381	-3.685	1.252834	-2.94	0.003	-6.144992	-1.225007
392	-2.925077	1.276064	-2.29	0.022	-5.430683	-.4194707
410	9.441614	1.257458	7.51	0.000	6.972541	11.91069
484	-5.280707	1.820292	-2.90	0.004	-8.854929	-1.706486
528	-3.06649	1.637391	-1.87	0.062	-6.281578	.1485983
579	-2.800006	1.359734	-2.06	0.040	-5.469901	-.1301106
616	1.581486	1.861233	0.85	0.396	-2.073125	5.236096
620	-3.040157	1.405664	-2.16	0.031	-5.800238	-.2800757
702	4.634749	1.679904	2.76	0.006	1.336184	7.933314
724	-1.828315	1.419122	-1.29	0.198	-4.614823	.9581924
752	-3.213421	1.419175	-2.26	0.024	-6.000031	-.4268099
757	-5.003736	2.220882	-2.25	0.025	-9.364534	-.6429394
764	-2.883901	1.686347	-1.71	0.088	-6.195116	.4273134
792	.6773891	1.535019	0.44	0.659	-2.336687	3.691465
826	-2.255673	1.368378	-1.65	0.100	-4.942542	.4311956
842	-1.777092	1.250489	-1.42	0.156	-4.232481	.6782966
_cons	11.25337	.9932716	11.33	0.000	9.303036	13.2037

.

c) **Compute the average for logGDP_future and private_credit_past for each country over time. Create two new variables, where you subtract the mean from each variable. Run a regression of the demeaned logGDP_future on the demeaned private_credit_past. Compare the coefficient obtained and the R 2 measures to the two previous regressions.**

We get exactly the same coefficients as in the previous 2 questions. However, we get lower R-squared compared with the previous ones. More specifically, in this method of including fixed effects, R-squared decreases to 8.7% from 38.7%. The table is shown below;

Source	SS	df	MS	Number of obs	=	695
				F(1, 693)	=	65.96
Model	1744.89659	1	1744.89659	Prob > F	=	0.0000
Residual	18332.2542	693	26.4534693	R-squared	=	0.0869
				Adj R-squared	=	0.0856
Total	20077.1508	694	28.9296121	Root MSE	=	5.1433

logGDP_new	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
private_credit_new	-.1186336	.0146071	-8.12	0.000	-.147313	-.0899541
_cons	7.27e-09	.1950962	0.00	1.000	-.3830506	.3830506

d) Which is the appropriate R 2 measure to report and why?

We think R-squared in part c (with demeaned data) is more appropriate to use. It is because, in part b, R-squared may be higher just because we added new repressors (dummy variables), but in part c we did not include any additional repressors. The basic mechanism for this result is that R-squared always increases (or stay the same) as we add new repressors (relevant or irrelevant).

e) What is the key identifying assumption, if you want to give the coefficient on private_credit_past (obtained including country FE) a causal interpretation?

The assumption would be that there is no time-variant omitted variables for each country and also there is no time-fixed effects (e.g. common shocks to all countries across time).

f) Now, run the same regression without country fixed effects. What do you conclude when comparing the obtained coefficient to the FE regression coefficient? Refer to your answer in the previous question.

As shown below, this time we get different coefficient, higher than (in absolute value) previous model. This coefficient may be higher because of omitted variable bias, in our case time-invariant country fixed effects. That means there can be some factors specific to each country (time-invariant) that works in the background affecting both private credit and GDP growth, so higher coefficients may include the impact of that factor as well.

Source	SS	df	MS	Number of obs	=	695
				F(1, 693)	=	63.44
Model	2507.70319	1	2507.70319	Prob > F	=	0.0000
Residual	27394.3454	693	39.5300799	R-squared	=	0.0839
				Adj R-squared	=	0.0825
Total	29902.0486	694	43.0865253	Root MSE	=	6.2873

logGDP_future	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
private_credit_past	-.1249523	.0156881	-7.96	0.000	-.1557541	-.0941504
_cons	9.360369	.267168	35.04	0.000	8.835813	9.884925

g) In a seminar someone suggests controling for a country's net foreign assets in addition to the country fixed effects. What do you answer?

We think country fixed effects already includes this information as long as net foreign assets are time-invariant. However, if net foreign assets change substantially across time, then we can include it.

REGRESSION SET 2

a) Run the regression of logGDP_future on private_credit_past without country fixed effects and store the residuals.

We run the simple regression and used predict function to store residuals in the variable, named resid_GDP.

b) Using the formula shown in class (the Moulton factor), compute the ratio of

v(βˆ)cluster / v(βˆ)robust

where "cluster" means clustering by country. Hint: the average number of observations per country is 26.17, the variance of the number of observations per country is 159.52, and the intra-class correlation coefficient can be computed using loneway

- the variance of the number of observations per country :159.52
- # of observations per country: 26.17
- intra-class correlation of X variable: 0.2
- intra-class correlation of Y variable: 0.31
- Moultan factor: **1.71**

c) Now run the regression in a) again, clustering s.e. by country. Compare the actual ratio of s.e. to the one predicted in question b).

v(βˆ)cluster=0.301

v(βˆ)robust=0.166

The ratio is 1.8

The actual ratio is a bit higher than the one predicted by the Moultan factor.

REGRESSION SET 3

a) What are the different possibilities to cluster s.e. in this setting? Which is the most stringent, which is the least stringent option?

The first option: add country fixed effects and cluster standard errors by the Country code.

The second option: add country fixed effects and cluster standard errors by the Year.

We think the first option would be the most stringent because this option takes into account some persistent shocks happening for each country. However, the second option will not consider time variant shocks for each country, but only time-variant worldwide common shocks happening for all countries.

b) Run the regression of logGDP_future on private_credit_past with country fixed effects. Cluster s.e. once using the least stringent, once using the most stringent option, and compare both of them to s.e. without clustering.

First option: country FE and clustering by country

(Std. err. adjusted for 30 clusters in CountryCode)

logGDP_future	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
private_cr~st	-.1186336	.0296982	-3.99	0.000	-.1793732	-.0578939
_cons	9.311867	.2279598	40.85	0.000	8.845637	9.778098

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
CountryCode	30	30	0	*

Second option: country FE and clustering by Year

(Std. err. adjusted for 46 clusters in year)

logGDP_future	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
private_cr~st	-.1186336	.0202456	-5.86	0.000	-.1594102	-.0778569
_cons	9.311867	.5292135	17.60	0.000	8.245977	10.37776

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
CountryCode	30	0	30

c) Does the coefficient change across the different versions?

No, it should not change, because clustering standard errors does not affect the value of the coefficients, rather it affects their significance.

d) In a seminar, someone suggests that you do not have to cluster s.e. when including country FE. What do you answer?

We think it may not be right suggestion. Including fixed effects consider only time-invariant differences across groups, but not the time-variant factors. So by clustering standard errors we could take into account time-variant factors affecting individuals within the group.

REGRESSION SET 4

a) Please replicate columns (2), (3) and (4) of Table III in [Mian et al., 2017]

In all there regressions, country fixed effects are added and standard errors are clustered by country code.

Column 2

logGDP_future	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
household_~st	-.3656467	.0691307	-5.29	0.000	-.5070348	-.2242586
_cons	10.00465	.303145	33.00	0.000	9.384647	10.62465

Column 3

logGDP_future	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
firm_credit~t	-.0978232	.0362572	-2.70	0.012	-.1719776	-.0236688
_cons	8.718426	.1175581	74.16	0.000	8.477993	8.95886

Column 4

logGDP_future	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
household_~st	-.3369889	.0675281	-4.99	0.000	-.4750993	-.1988785
firm_credit~t	-.0410688	.0328202	-1.25	0.221	-.1081936	.0260561
_cons	10.01214	.2920518	34.28	0.000	9.414827	10.60945

b) Test formally whether the coefficients on household credit and firm credit in the regression of column (4) are equal. Provide the distribution, H0, HA and the result of the test.

```
. test household_credit_past== firm_credit_past

 ( 1)  household_credit_past - firm_credit_past = 0

       F(  1,    29) =   13.26
            Prob > F =    0.0010
```

H0: 2 coefficients are equal

HA: 2 coefficients are not equal

According to the test result, we can reject null hypothesis, so these 2 coefficients are not equal.

c) Re-run the regression of column (4) and add year fixed effects. Briefly comment on the results. How does the interpretation of the coefficients change compared to the regression without year fixed effects?

logGDP_future	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
household_~st	-.2214482	.0693343	-3.19	0.003	-.3632527	-.0796436
firm_credit~t	-.0378854	.0298214	-1.27	0.214	-.098877	.0231062
_cons	9.49516	.3247572	29.24	0.000	8.830957	10.15936

Adding year fixed effects (additional to country fixed effects) decreases the absolute value of the coefficients. That is expected, because now we can also control for common shocks happening worldwide to all countries across time.

5. QUESTIONS

a) Some countries in the dataset are quite close to each other geographically and likely to be integrated economically. Discuss the potential implications for the analysis in question 4). Expected length: about 6-8 sentences.

In this analysis, standard errors are clustered by countries, so it means that it is assumed that correlation of error terms across countries is zero. However, if the dataset contains some countries integrated each other economically then this assumption may be violated. That means, there would be some correlation across clusters as well. In that case, the standard errors may be underestimated (if the correlation of error terms across countries is positive), which may put doubt on the significance of the paper's findings.

```
use "Z:\Desktop\dataEmpBF_Tutorial4.dta"
*** QUESTION 1
ssc install reghdfe
ssc install ftools
reghdfe logGDP_future private_credit_past, absorb (CountryCode)
drop if logGDP_future == .
reg logGDP_future private_credit_past i.CountryCode
egen mean_logGDP=mean( logGDP_future), by(CountryCode)
egen mean_private_credit=mean( private_credit_past), by( CountryCode)
gen logGDP_new= logGDP_future- mean_logGDP
gen private_credit_new= private_credit_past- mean_private_credit
reg logGDP_new private_credit_new
reg logGDP_future private_credit_past

***QUESTION 2
predict resid_GDP, residuals
loneway logGDP_future CountryCode
loneway private_credit_past CountryCode
reg logGDP_future private_credit_past, robust
reg logGDP_future private_credit_past, cluster ( CountryCode)

***QUESTION 3
reghdfe logGDP_future private_credit_past, absorb( CountryCode) cluster ( CountryCode)
reghdfe logGDP_future private_credit_past, absorb( CountryCode) cluster ( year)

***QUESTION 4
reghdfe logGDP_future household_credit_past ,absorb( CountryCode) cluster(CountryCode)
reghdfe logGDP_future firm_credit_past ,absorb( CountryCode) cluster( CountryCode)
reghdfe logGDP_future household_credit_past firm_credit_past ,absorb( CountryCode) cluster(
CountryCode )
test household_credit_past== firm_credit_past
reghdfe logGDP_future household_credit_past firm_credit_past, absorb( CountryCode year)
cluster( CountryCode)
```