# Empirical Banking and Finance: Tutorial № 4

Altaf Hussain and Simon Bong 05.06.2020

## Set WD, Install Packages and Load Data

```
cd "Z:\Downloads"

ssc install estout, replace
ssc install reghdfe, replace
ssc install ftools, replace

use dataEmpBF_Tutorial4.dta
```

## 1 Data & Descriptives

### 1.1 Create descriptive statistics and comment briefly.

```
estpost summarize  // should delete observations with NAs
esttab using sumr.tex, cells("count mean sd min max") noobs
```

Table 1: Summary Statistics for all Variables

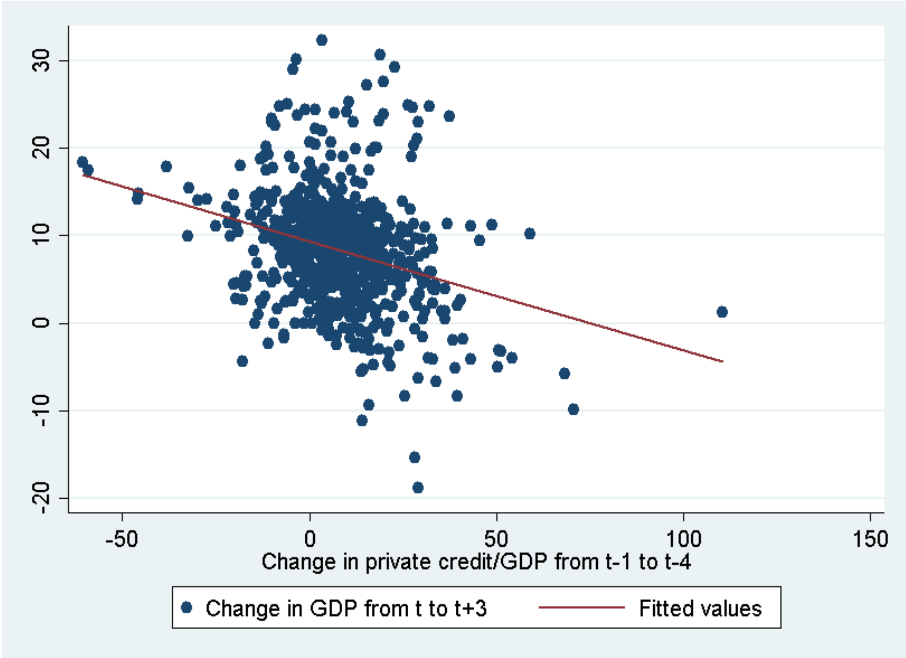
| | count | mean | sd | min | max |
|---|---|---|---|---|---|
| CountryCode | 785 | 445.0994 | 250.5633 | 36 | 842 |
| c | 0 | . | . | . | . |
| year | 785 | 1996.469 | 11.55728 | 1964 | 2012 |
| logGDP | 785 | 28.32083 | 2.400927 | 24.64919 | 35.50195 |
| logGDP_future | 695 | 8.401251 | 6.564033 | -18.8221 | 32.32899 |
| private_credit | 785 | 128.2249 | 54.83109 | 17.02012 | 339.2 |
| household_credit | 785 | 49.92093 | 28.04941 | .1457143 | 144.7189 |
| firm_credit | 785 | 78.80823 | 33.25036 | 7.479578 | 227.4264 |
| private_credit_past | 785 | 8.677822 | 16.09333 | -60.47037 | 114.603 |
| household_credit_past | 785 | 4.60116 | 6.189944 | -11.19909 | 34.95463 |
| firm_credit_past | 785 | 4.016322 | 12.37292 | -48.23518 | 86.50771 |

- CountryCode/c: We see that the Country has no missing values. Looking on other statistics makes no sense since the variable is just an abitrary ID.
- Year: We have no missing values for any observations and observations are between the years 1964 and 2012.
- LogGDP: The log for the real GDP in constant prices has also no missing obervations. We see values between 24.6 and 35.5, the interpretaion is not really possible since the values are messured in local currency units.
- LogGDP_future: There are 90 missing values. For some countries the change in GDP from t to t+3 is negative and for other it is poitive.
- Private_Credit: Private non-financial debt to GDP varies a lot, the minimum value is 17 while the maximum value is 339.
- household_credit: Household debt to GDP also varies a lot, from nearly zero up to 144.
- firm_credit: Non-financial firm debt to GDP also varies a lot, from 7 to 227.
- private_credit_past[1]: The change in the ratio of private credit to GDP from t-4 to t-1 is in mean about 8 and has a range from -60 to 114.
- household_credit_past: The change in the ratio of private household to GDP from t-1 to t-4 is in mean about 4 and has a range from -11 to 35.
- Firm_Credit_Past: The change in the ratio of firm credit to GDP from t-1 to t-4 is in mean about 4 and has a range from -48 to 87.

### 1.2 Correlation of past change in private credit and future change in GDP

```
twoway scatter logGDP_future private_credit_past ///
|| lfit logGDP_future private_credit_past
corr logGDP_future private_credit_past
```

[1] according to the label of variable in the dataset, the change is mesured form t-1 to t-4. But since its counterintuitve and also diffierent in the Milan (2017)




Correlation coefficient = -0.2896

- logGDP_future: For each point in time the difference of the GDP three years later and the GDP of that year is calculated. The real GDP is used.
- private_credit_past: For each point in time the difference of the private credit to GDP of the previous year and the same measure from four years earlier is calculated.

We see a negative correlation of the two variables, implying a negative relation between the change in private credit to GDP and subsequent output growth. The relation is not perfect ($|r| \neq 1$) and medium strong.
A theoretical explanation could be that banks are most willing to lend to households when unemployment is low and the ratings of their customers are good and vice versa, when we assume the demand function does not change over time this implies a pro-cyclical amount of debt over time. When the business cycle has a length of around six years this theory fits to the data and is also consistent with the longrun findings (trend) in the papers of the previous Problem Sets.

### 1.3 How does the result fit to results we have seen in previous Problems

Porta et al (1998) and Levine et al (2000) both find a positve relation between private credit to GDP and subsequent growth. The findigs are not necessary inconsistent with those of Milan (2017) since Milan (2017) considers the short run while the other two papers look on a longer time horizon. So it might be the case that the relation is pro-cyclical and the trend of growth is positively influenced by private credit at the same time.

## 2 Regression set 1

### 2.1 Regression of future LogGDP on past private Credit with country fixed effects

```
reghdfe logGDP_future private_credit_past, a(c)
est sto m1
esttab m1 using reg1.tex, se obslast scalar(F) r2 ///
title("Regression 1") replace
```

We observe a negative relation, when private_credit_past increases by one standard deviation in average logGDP_future is 0.2918 standard deviations lower:

Table 2: Regression 1

| Model | $y_{i,t} = \alpha_i + \beta x_{i,t} + \varepsilon_{i,t}$ |
|---|---|
| | logGDP_future |
| private_credit_past | -0.119*** |
| | (0.0149) |
| _cons | 9.312*** |
| | (0.230) |
| $R^2$ | 0.387 |
| F | 63.20 |
| $N$ | 695 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 2.2 Regression 1 with reg command

```
reg logGDP_future private_credit_past i.CountryCode, robust
est sto m2
esttab m2 using reg2.tex, se obslast scalar(F) r2 ///
title("Regression 2") replace
```

The results for the coefficent as well as for the $R^2$ are similar as expected, due to the fact that the methods of fixed effects and first differences are similar in terms of the coefficents of interest:

Table 3: Regression 2

| Model | $y_{i,t} = \alpha_i + \beta x_{i,t} + \varepsilon_{i,t}$ |
|---|---|
| | logGDP_future |
| private_credit_past | -0.119*** |
| | (0.0163) |
| _cons | 11.25*** |
| | (0.472) |
| $R^2$ | 0.387 |
| F | 34.08 |
| $N$ | 695 |

Standard errors in parentheses

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

### 2.3 Regression 1 using demeaned variables

```
by c, sort : summarize logGDP_future private_credit_past
foreach i in logGDP_future private_credit_past{
egen mean_‘i’=mean(‘i’), by(c)
ge demeaned_‘i’=‘i’-mean_‘i’
}
reg demeaned_logGDP_future demeaned_private_credit_past, robust
est sto m3
esttab m3 using reg3.tex, se aic obslast scalar(F) bic r2 ///
title("Regression 3") replace
```

Again we observe the same[2] coefficient, because the methods are same[3] here. The $R^2$ differs as we see:

[2] the slight difference is due to computation errors/rounding in the demeaning process
[3] in terms of the estimation of the coefficent of the indipendent variable

Table 4: Regression 3

| Model | $y_{i,t} - \bar{y}_i = \beta(x_{i,t} - \bar{x}_i) + \varepsilon_{i,t}$ |
|---|---|
| | demeaned_logGDP_future |
| demeaned_private_credit_past | -0.117$^{***}$ |
| | (0.0156) |
| _cons | -0.0891 |
| | (0.197) |
| $R^2$ | 0.086 |
| F | 56.82 |
| $N$ | 695 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 2.4 Which $R^2$ should we report

We should report the $R^2$ from the demeaned ols regression, because it gives us the most adequate measure of the goodness of fit imposed by the variable of interest. In the other specifications the $R^2$ is much higher and dependent on fixed effects variables due to the explainatory power[4] of them.

### 2.5 Assumption allowing causal interpretation

We have to assume that there is no omitted variable[5], which is not constant overtime for each given country. Then we would have the normal OVB problem, if the omitted variable is constant overtime for each given country we would already control for it by the country fixed effects.

### 2.6 Fixed effects for countries legal origin, a good idea?

When clustering over country the clustering over legal origin is already controlled for - the legal origin of a country is constant over time.

[4] not neccessary casual power, but residual variance is reduced

[5] which is correlated with the residual and the indipendent variable

# 3 Regression set 2

## 3.1 Regression of logGDP_future on private_credit_past with country fixed effects

- (1) Using the default standard errors
- (2) Using standard errors robust to hetereoskedasticity
- (3) Clustering standard errors at the country level
- (4) Clustering standard errors at the country level and at the year level

```
reghdfe logGDP_future private_credit_past, a(c) //1
est sto d1
reghdfe logGDP_future private_credit_past, a(c) vce(r) //2
est sto d2
reghdfe logGDP_future private_credit_past, a(c) vce(cluster c) //3
est sto d3
reghdfe logGDP_future private_credit_past,a(c) vce(cluster c year) //4
est sto d4
esttab d1 d2 d3 d4 using reg4.tex, se obslast scalar(F) r2 ///
title("Regression 4") replace
```

Table 5: Regression 4

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | logGDP_future | logGDP_future | logGDP_future | logGDP_future |
| private_credit_past | -0.119$^{***}$ | -0.119$^{***}$ | -0.119$^{***}$ | -0.119$^{***}$ |
| | (0.0149) | (0.0163) | (0.0297) | (0.0319) |
| _cons | 9.312$^{***}$ | 9.312$^{***}$ | 9.312$^{***}$ | 9.312$^{***}$ |
| | (0.230) | (0.219) | (0.228) | (0.526) |
| $R^2$ | 0.387 | 0.387 | 0.387 | 0.387 |
| F | 63.20 | 52.99 | 15.96 | 13.85 |
| $N$ | 695 | 695 | 695 | 695 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 3.2 Compare the standard errors across the four types of adjustments and comment

We see that the standard error assuming homoscedasticity is lowest and the robust standard error is higher when we cluster over countries. Clustering additionaly over years increases the standard error even more.
As we already know from the very beginning of the lecture robust standard errors are always higher than those calculated assuming homoscedasticity. We also know that due to the additional covariance terms usually clustered standard errors are even higher than simple robust standard errors. The clustered standard errors for country and year are higher than those clustered only over countries, since the number of observations available for the estimation of each element of the covariance matrix is lower with more clusters.

### 3.3 What are the "theoretical" arguments in favor of and against clustering standard errors at the country-level? What could be the reason for the additional clustering at the year level?

- Main argument for clustering: If we do not cluster standard errors / assume that observations are iid while this is not the case we underestimate the standard errors by a large extent and maybe wrongly report significance of our coeficients.
- (Argument) against clustering: If we cluster standard errors while the iid assumption holds we overestimate the standard errors, but this is hard to argue.
- Argument for clustering over Years: Clustering over years makes much sense since its commonly known and empirical proven that there is often autocorrelation to a large extent for GDP variables.

### 3.4 Does the coefficient change?

No, when we look on their calculation than we see that the covariance matrix is not part of the equation.

### 3.5 Clustering standard errors at the country-year level, a good idea?

The result would be the same covariance matrix as with normal robust estimation. So since we want to check for the covariances and not underestimate our standard errors, it is a bad idea.

## 4 Regression set 4

### 4.1 Replication of columns (2), (3) and (4) of Table III in [Mian et al., 2017]

```
reghdfe logGDP_future household_credit_past , ///
a(c) vce(cluster c year) // HH
```

```
est sto f1
reghdfe logGDP_future firm_credit_past, ///
a(c) vce(cluster c year) // Firm
est sto f2
reghdfe logGDP_future household_credit_past firm_credit_past, ///
a(c) vce(cluster c year) // Firm and HH
est sto f3
reghdfe logGDP_future household_credit_past firm_credit_past, ///
a(c year) vce(cluster c year) // Firm and HH and additional year FE
est sto f4
esttab f1 f2 f3 f4 using reg5.tex, se obslast scalar(F) r2 ///
title("Regression 5") replace
```

Table 6: Regression 5

| | (2) | (3) | (4) | (4 add year FE) |
|---|---|---|---|---|
| | logGDP_future | logGDP_future | logGDP_future | logGDP_future |
| household_credit_past | -0.366*** | | -0.337*** | -0.221** |
| | (0.0786) | | (0.0794) | (0.0674) |
| firm_credit_past | | -0.0978* | -0.0411 | -0.0379 |
| | | (0.0398) | (0.0355) | (0.0324) |
| _cons | 10.00*** | 8.718*** | 10.01*** | 9.495*** |
| | (0.463) | (0.560) | (0.452) | (0.319) |
| $R^2$ | 0.411 | 0.353 | 0.415 | 0.659 |
| F | 21.64 | 6.038 | 12.30 | 5.605 |
| $N$ | 695 | 695 | 695 | 695 |

Standard errors in parentheses

$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$

### 4.2 F-Test: Are the coefficients on household credit and firm credit in the regression of column (4) are equal

```
reghdfe logGDP_future household_credit_past firm_credit_past, ///
a(c) vce(cluster c year) // Firm and HH
test firm_credit_past==household_credit_past
```

- $H_0$: The coefficients are equal $\iff \beta_{HHCredPast} - \beta_{FirmCredPast} = 0$
- $H_1$: Negation of $H_0$
- The Teststatistic is F Distributed: $F(1, 29) = 9.45\ Prob > F = 0.0046$

- We reject the $H_0$, that the coefficients are equal, at the 1% level

### 4.3 Re-run the regression of column (4) and add year fixed effects

```
reghdfe logGDP_future household_credit_past firm_credit_past, ///
a(c year) vce(cluster c year) // Firm and HH and additional year FE
```

The results are already shown in table 6 column (4 add year FE).
We see that the $R^2$ went up by a large extend, but this was expected and does not tell us much. Standard errors where not effected by large extend. Also all relevant coeficients did not change by significant extent except for the coefficent of household_credit_past. This coeficient rised from -0.337 to -0.221 which implies that there probably was an omitted variable correlated with the error of the regression and household_credit_past, which is constant accross households but changes over time. This caused a downward biased coefficent.
Example: a nonpersistent shock in t-1 hits all households and varies over time, it is negatively correlated with household_credit_past and has an positve impact on the change of GDP from t to t+3 by reducing t but not t+3.