Empirical Banking and Finance: Tutorial № 2

Altaf Hussain and Simon Bong 01.05.2020

1 Regression 1

1.1 Load Dataset

```
1 use dataEmpBF_Tutorial2.dta
```

1.2 OLS regression of gdpgrowth on public_banks_1970 with loggdp_1960 as a control

```
1 reg gdpgrowth public_banks_1970 loggdp_1960, robust
```

```
Linear regression                               Number of obs     =         86
                                                F(2, 83)          =       0.71
                                                Prob > F          =     0.4948
                                                R-squared         =     0.0557
                                                Root MSE          =     .03377
```

gdpgrowth	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
public_~1970	.0192503	.030972	0.62	0.536	-.0423518	.0808523
loggdp_1960	-.0007913	.0053465	-0.15	0.883	-.0114253	.0098428
_cons	.0133436	.0447428	0.30	0.766	-.075648	.1023352

The model as a whole is not significant (F statistic) as well as the single coefficents (t-test).

1.3 Compare the estimated coefficient of public_banks_1970 to the first row in Table V of [La Porta et al., 2002]

La Porta et al are reporting a negative significant coefficient (-0.0235) for GP70 while we estimate a positive and insignificant coefficient. Their result indicates that a 1% higher share of goverment ownership of banks results in average in a 2.35% lower (subsequent) per capita gdp growth. This supports the political theory.
Besides the coefficent for the starting gdp level indicates possible concergence effects.

2 Descriptive Statistics

2.1 Computing summary statistics for all variables

```
1  su
```

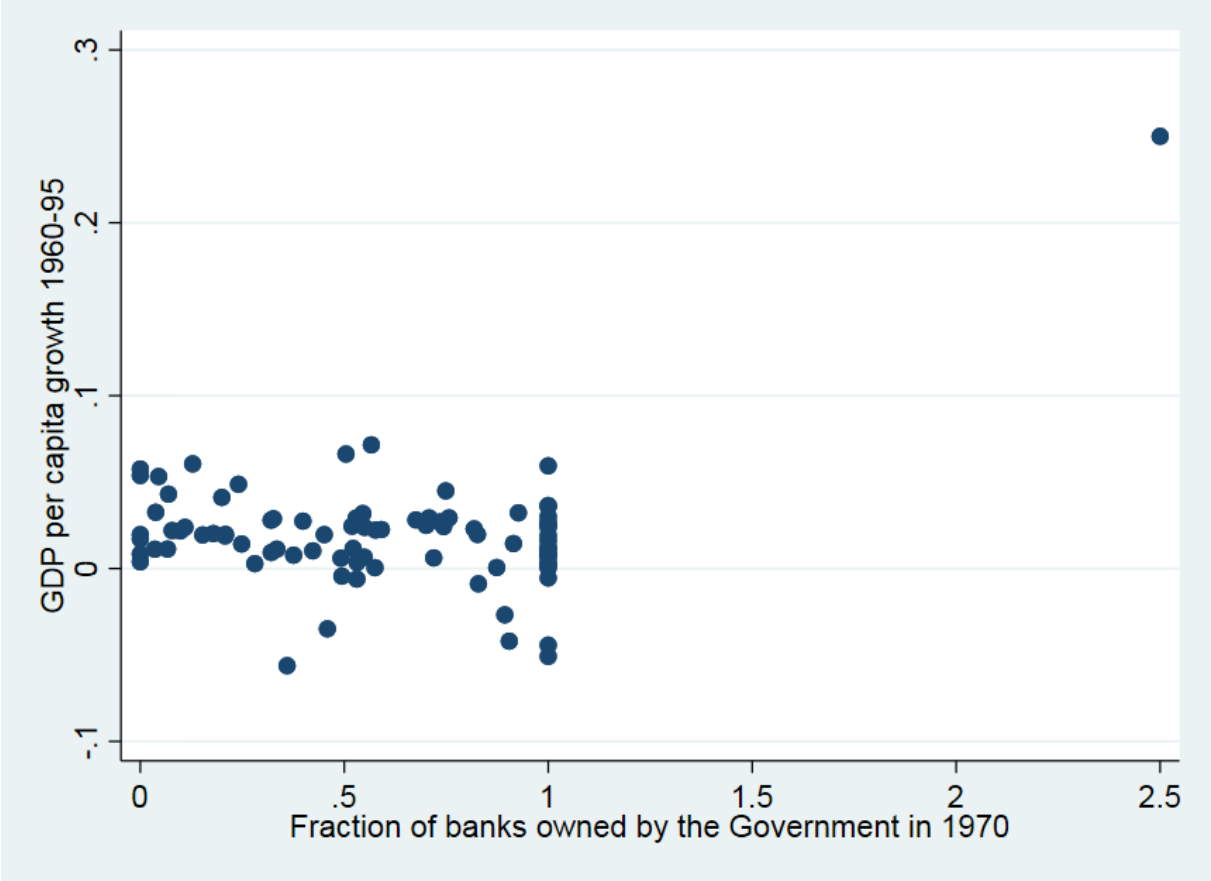
Variable	Obs	Mean	Std. Dev.	Min	Max
country_name	0				
birth_r~1970	83	32.84936	12.94351	13.5	53.32
public_~1970	86	.5839216	.4056251	0	2.5
schooling	85	5.030555	2.459215	.9	10.79
private~1960	82	.254575	.2294715	.0166615	1.295776
gdpgrowth	86	.0199698	.0343439	-.0561875	.25
loggdp_1960	86	5.83181	1.00547	3.931826	8.670402
oecd	85	.2823529	.4528157	0	1

Over all we see that the number of observations differs across the variables, this indicates missing values for some variables

- Birth rate per 1000 population in 1970: When we look at the standard deviation we see that there are big differences across countries
- Fraction of banks owned by the Government in 1970: most imortant, there is at least one observation with a value over 100% which is not possible (max=250%)in avg the goverment of a country owns 58% of the assets of the top 10 banks in that country.
- Average years of schooling 1960-90: in average childs go 5years to school and the range of the observations goes from less then one year up to than 10.8
- Private credit to GDP in 1960: this share is in mean equal to 25%. The cross country differences are huge, what is indicated by min max and sd
- GDP per capita growth 1960-95: we see in average there was a small per capita in gdp growth, but again we see huge cross country differences.
- log GDP per capita in 1960: here we also see that the level of the gdp per capita in 1960 differs alot across countries.
- Old OECD member states: we see that 28% of the countrys are old OECD member states (when we just include those with non missing values)

2.2 Scatter Plot gdp per capita growth 1960-1995 and GP70

```
1 scatter gdpgrowth public_banks_1970
```



2.3 Remove outlier

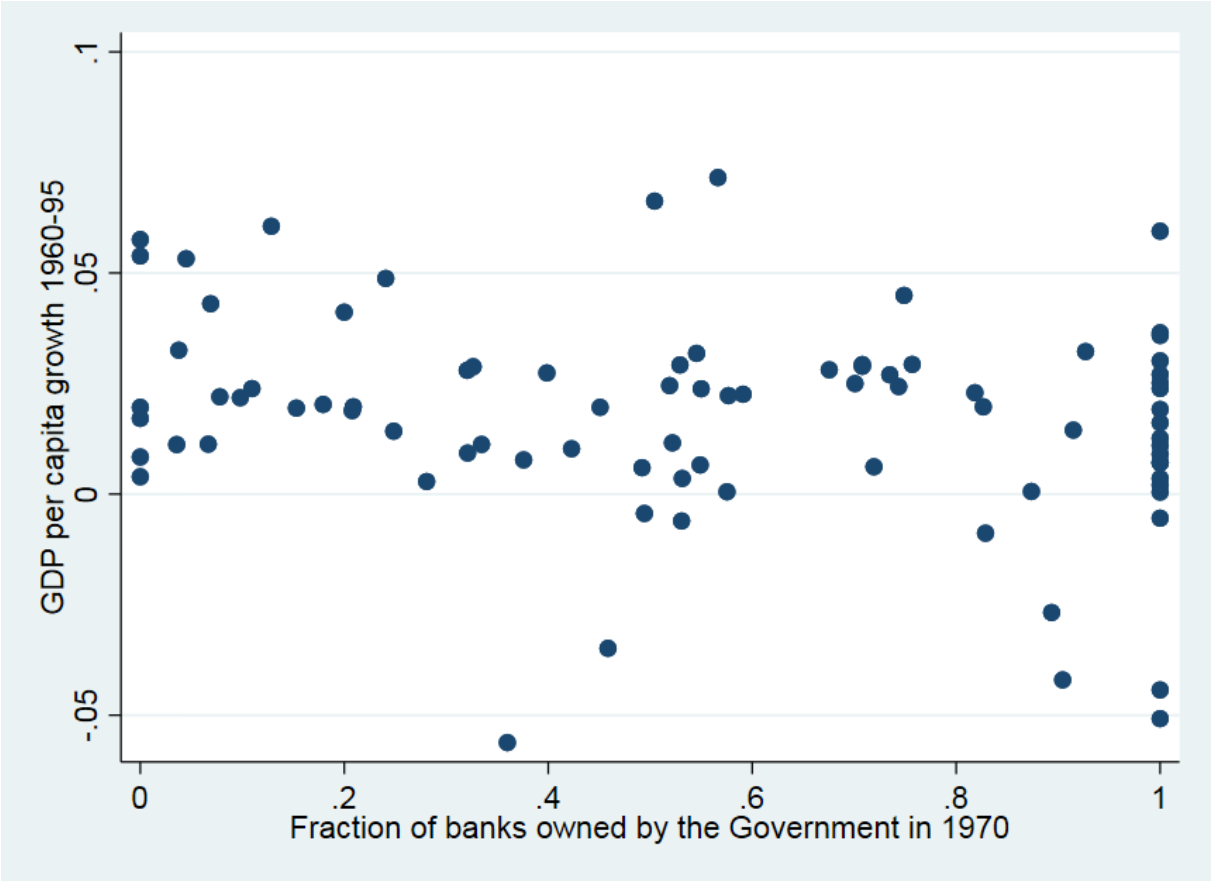
We can savely remove the county with a public_banks_1970 value of 2.5 because a goverment can not own more then 100% of the assets of the 10 greatest banks.

```
1 drop if public_banks_1970>1
```

We see that one observation was deleted.

2.4 Scatter Plot gdp per capita growth 1960-1995 and GP70 w/o outlier

```
1 scatter gdpgrowth public_banks_1970
```



3 Regression 2

3.1 OLS regression of gdpgrowth on public_banks_1970 with loggdp_1960 as a control w/o outlier

```
1  reg gdpgrowth public_banks_1970 loggdp_1960, robust
```

```
Linear regression                               Number of obs     =         85
                                                F(2, 82)          =       4.93
                                                Prob > F          =     0.0095
                                                R-squared         =     0.1240
                                                Root MSE          =    .02234
```

gdpgrowth	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
public_~1970	-.023484	.0076929	-3.05	0.003	-.0387877	-.0081804
loggdp_1960	-.0064537	.0031549	-2.05	0.044	-.0127297	-.0001776
_cons	.0680836	.0205328	3.32	0.001	.0272374	.1089298

Now after we deleted the outlier you added, we observe we have the same results as in Table V of La Porta.

OLS is very sensetive to outliers since it is minimizing the (empirical) MSE. Observations 'far away' from 'the rest', therefore get a overproportional weight.

3.2 Why using 1970 Data?

Why are the authors using government ownership of banks in 1970 and not, for example average ownership share between 1970 and 1995?
One the one hand they report a relatively high correlation over time (eg .9 for GP70 and GP 85) and on the other hand they argue with the availability of the data. Besides they report that the results using GP95 are very simmilar.

Would it be preferable to use government ownership in 1960, and if yes, why?
Yes, this way we can be sure that the goverment ownership is not effected by the dependend variable (exclude the possibility that the goverment ownership level is high because of low gdp, banks failed goverments bought assets to save them or some other reverse relationship). This way one could say X is predetermined.

3.3 Significance

Yes, the coefficent for GP75 is higly significant (p=0.003) and the coefficent for loggdp960 is significant on the 5% level (p=0.044).

3.4 Interpretation of coefficents

looking on the individual effects:
As the share of goverment ownership of banks increases, in average the gdp growth per capita goes down. A 1% higher goverment ownersip in 1970 results in a 2.484% lower gdp growth by capita between 1960-1995.
The negative coefficent for the gdp starting level indicates possible convegence effects.
A 1% higher starting gdp per capita level results in avg a 0.645% lower gdp per capita growth from 1960-1995.

3.5 Model Fit

Our X Variables explain 12.4% of the varince of the depended variable (see R^2) and the model as a whole is significant (see F statistic). Assuming that the true coefficents are all equal to zero, observing this F value or a more extreme one has a probability of less then 1%.

4 Regression 3

4.1 avoid Omitted Variable Bias

Calculate Correlation Matrix

```
1 corr
```

	bir~1970	pub~1970	school~g	pri~1960	gdpgro~h	log~1960	oecd
birth_r~1970	1.0000						
public_~1970	0.2595	1.0000					
schooling	-0.8608	-0.3636	1.0000				
private~1960	-0.5406	-0.2479	0.4509	1.0000			
gdpgrowth	-0.3934	-0.2482	0.2681	0.3550	1.0000		
loggdp_1960	-0.6706	-0.3979	0.6965	0.3773	-0.1325	1.0000	
oecd	-0.7244	-0.2681	0.6412	0.4648	0.2941	0.5451	1.0000

We have a potential problem with OVB since we see a non zero cerrelation of the two variables (we are going to include) and GP70 and simultainosly most certainly there is some relation between the two variables and our dependent variable.

```
1 reg gdpgrowth public_banks_1970 loggdp_1960 schooling birth_rate_1970, ↩
    robust
```

```
Linear regression                               Number of obs     =         83
                                                F(4, 78)          =      24.60
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5390
                                                Root MSE          =     .01614
```

gdpgrowth	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
public_~1970	-.0263877	.0064036	-4.12	0.000	-.0391362	-.0136391
loggdp_1960	-.0191741	.002353	-8.15	0.000	-.0238586	-.0144897
schooling	-.0009434	.0012879	-0.73	0.466	-.0035073	.0016206
birth_r~1970	-.0016713	.0002268	-7.37	0.000	-.0021228	-.0012199
_cons	.2036386	.020668	9.85	0.000	.1624917	.2447854

4.2 Significance

All variables are higly significant exept for schooling.

4.3 Interpretation

A higher share of goverment ownership of banks results in lower gdp per capita growth in the future. That supports the political theory.
A higher starting point of gdp per capita results in lower subsequent growth, like it is predicted by the theory of convergence effects.
A higher birthrate also causes a loger gdp per capita growth. This supports that in the considered time window the negative effects of a increasing denumerator of gdp per capita and the temporary lower labor force due to taking care for the children outwights the effect of more furute labor force.

4.4 Test for the joint significance of schooling and birth_rate_1970

- H0: both variables have a true coefficent equal to zero
- HA: at least one coefficent is ne 0

We use a F test, the F-Statistic is follows a F distribution (when all assumptions hold)
Basicly we compare the R^2 of a regression with and without the two variables.

```
1 test schooling birth_rate_1970
```

```
schooling = 0
birth_rate_1970 = 0

F(  2,     78) =    40.94
      Prob > F =     0.0000
```

The result is that we reject the H0 at the 1% lecel and that is it very unlikely that the true coefficents simultaneously equal to zero.

4.5 Compare Goodness of Fit to Regression 2

One option is to compare the R^2 for both regressions, but this we already did in the previous subtask by doing the F-Test. If we want to take account for the increase in dimension we could look ont the adj R^2.

5 Regression 4

5.1 Why look on interaction of GP70 with private credit

Goverment owned banks are maybe not efficent when they are the only lender in the financial system, but they are maybe a efficent comlement to a well working financal market in the sense that they can step in when market failtures occur.
The private credit to gdp can be used as a mesurement for the development of the financial market.

5.2 Run Regression whith interactionterm GP70*private credit

```
1 reg gdpgrowth c.public_banks_1970##c.private_credit_1960 loggdp_1960 , ↩
    robust
```

Linear regression

Number of obs	= 82
F(4, 77)	= 7.70
Prob > F	= 0.0000
R-squared	= 0.3159
Root MSE	= .02035

gdpgrowth	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
public_~1970	-.0397587	.0106868	-3.72	0.000	-.0610389	-.0184784
private~1960	.0186234	.0107444	1.73	0.087	-.0027714	.0400181
c. public_~1970# c. private~1960	.0685466	.0194724	3.52	0.001	.029772	.1073211
loggdp_1960	-.0107444	.0029275	-3.67	0.000	-.0165739	-.004915
_cons	.0892572	.019143	4.66	0.000	.0511386	.1273757

5.3 Interpretation the coefficents for private credit and GP70

When a country has zero private credit, then a 1% higher share of goverment ownership of banks results in avg in a 3.97% decrease in the subsequent gdp per capita growth.
when a country has zero goverment ownership of banks, then a 1% higher privatecredit of gdp share results in avg in a 1.86% increase in the subsequent gdp per capita growth.

5.4 Interpretaion of the sign of the interaction term

Here we can only give a qualitative interpretaion or make pointwise comparisons, meaning compare the Prediction for different X varaible values.
The higher the share of private credit to gdp, the better the marginal effect of public banks. The marginal effect is even positive given a value of private credit higher than 58% (but this isn't significant at that point).

5.5 Calculate margrinal effect of GP70

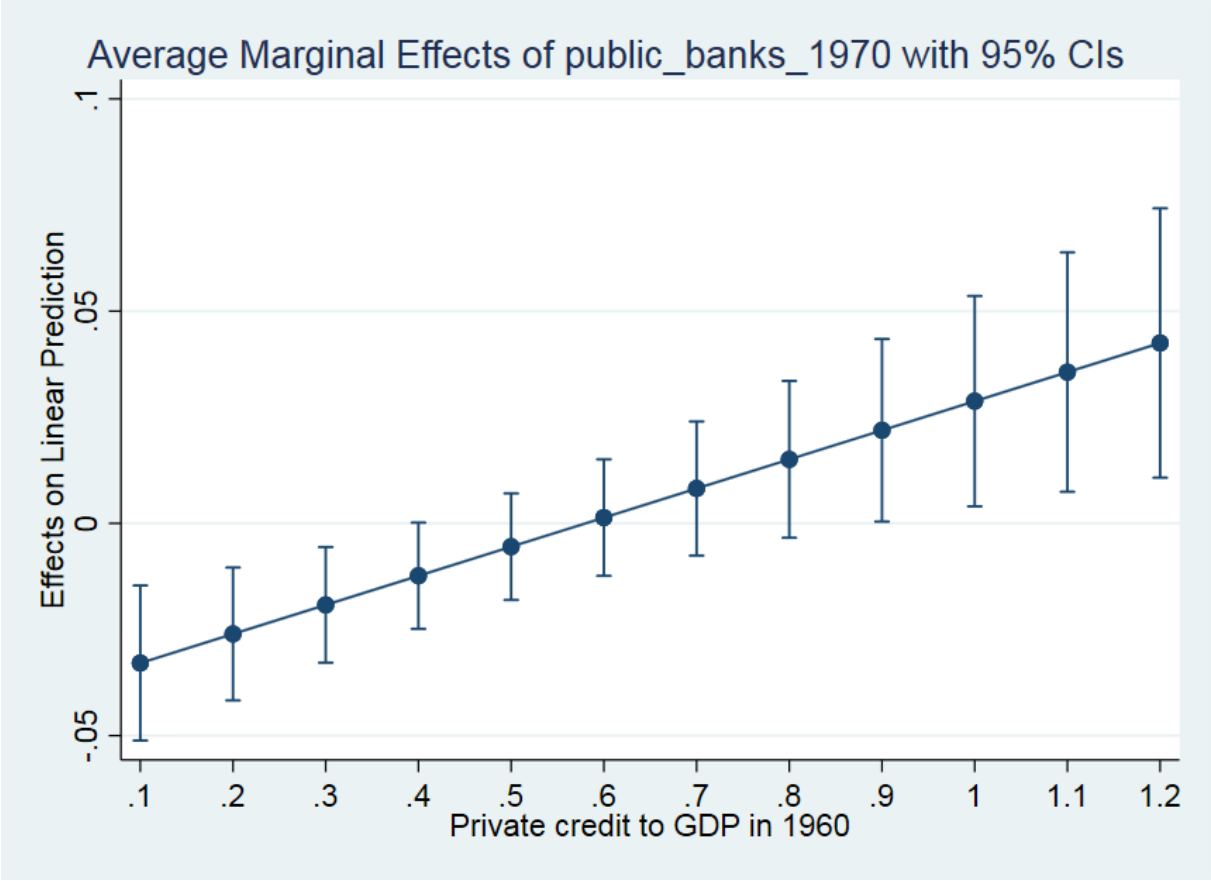
First we did it inefficent, then the normal way:

```
1 ge margin = -.0397587+.254575*.0685466
2 drop margin
3
4 margins, dydx(public_banks_1970) atmeans
```

-.0223085 is the marginal effect evaluatet at the mean of private credit (and all other vars).

5.6 Significance for different values of private credit

```
margins, dydx(public_banks_1970) at(private_credit_1960=(0.1 0.2 0.3 0.4 ↩
    0.5 0.6 0.7 0.8 0.9 1 1.1 1.2))
marginsplot, level(95)
```



The marginal effect of GP70 isn't significant in the range of approxximately 40% to 90% for private credit in 1960.

6 Regression 5

6.1 Run regression with dummyvariables and interaction

oecd or noecd will be exluded due to multicolinearity we, out of the set intercept oecd and noecd always one variabel can be expressed as a linear combination of the other two.

```
ge noecd = (oecd==0)
reg gdpgrowth noecd i.oecd##c.public_banks_1970 i.oecd##c.loggdp_1960
```

Source	SS	df	MS	Number of obs	=	85
				F(5, 79)	=	7.25
Model	.014695264	5	.002939053	Prob > F	=	0.0000
Residual	.032026362	79	.000405397	R-squared	=	0.3145
				Adj R-squared	=	0.2711
Total	.046721625	84	.00055621	Root MSE	=	.02013

gdpgrowth	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
noecd	-.0131905	.0498535	-0.26	0.792	-.1124214	.0860403
1.oecd	0	(omitted)				
public_~1970	-.0283474	.0078303	-3.62	0.001	-.0439333	-.0127616
oecd# c. public_~1970 1	.0264119	.0162471	1.63	0.108	-.0059272	.0587511
loggdp_1960	-.0120677	.0030166	-4.00	0.000	-.0180722	-.0060633
oecd# c. loggdp_1960 1	-.0001172	.0070915	-0.02	0.987	-.0142325	.0139982
_cons	.1101278	.0462736	2.38	0.020	.0180224	.2022332

6.2 Interpretation of the constant and the coefficent of the dummy

since gdp level in 1960 >0 it makes no sense to look on the constant seperately.
The coefficent for noecd dummy can not be interpreted seperately since the normal mean comparison makes no sense when we do not take account for our interaction terms. We should instead look on marginal effects.

6.3 Interpretation of the coefficents of the interactionterms

a clean interpretation of the interaction coefficents is also only pointwise possible. Due to similar reasons as in b).

7 Regression 6

7.1 Run regression with two 'intercepts'

```
1 reg gdpgrowth noecd i.oecd##c.public_banks_1970 i.oecd##c.loggdp_1960, ↩
    robust noconstant
```

Linear regression

Number of obs	85
F(6, 79)	84.43
Prob > F	0.0000
R-squared	0.5555
Root MSE	.02013

gdpgrowth	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
noecd	.0969373	.0217118	4.46	0.000	.053721	.1401536
1.oecd	.1101278	.034344	3.21	0.002	.0417678	.1784879
public_~1970	-.0283474	.0092809	-3.05	0.003	-.0468206	-.0098743
oecd# c. public_~1970 1	.0264119	.0105571	2.50	0.014	.0053986	.0474253
loggdp_1960	-.0120677	.0036771	-3.28	0.002	-.0193868	-.0047486
oecd# c. loggdp_1960 1	-.0001172	.0060282	-0.02	0.985	-.012116	.0118816

7.2 Interpretation of the two dummies

The coefficient for the dummy OECD is the intercept for the oecd observations and the coefficent for the dummy no-oecd is the intercept for the non OECD countries.
The coefficient of the intersection of public banks and oecd gives us the difference in the effect of a marginal increase in public banks between oecd and no oecd countries. one percent increase in the share of government owned banks increases per capita gdp growth by 2.64% more when the country is a old oecd member state compared to a non member state.
The coefficient of the intersection of loggdp_1960 and oecd gives us the the difference in the effect of a marginal increase in loggdp_1960 between oecd and no oecd countries, this coefficient is not significant (alone).
To summarize, we got the same results as if we run two regressions, one for oecd members and one for non-oecd members. But we put the results together in one model - the fitted values are exactly the same for our model as the two separate regressions.

7.3 Why do you have to exclude the constant?

We have two dummies where always exactly one of them is equal to one, so if we include a constant we would get trouble with co-linearity. The constant would be a linear combination of the two dummies.

7.4 Compare Results to Regression 5

In the previous regression, we had a problem with to colinearity of the dummies and the constant, which stata solved by omitting one dummy variable. But the version of the model with a dummy and a intercept(Regression 5) is not that clear to interpret as the one with two dummies (Regression 6).