Statistics/Data Analysis

```
      name:  <unnamed>
       log:  C:\Users\s3sibong\Downloads\Tutorial2.smcl
  log type:  smcl
 opened on:   2 May 2020, 13:58:51

1 . use dataEmpBF_Tutorial2.dta

2 .
3 .
4 . // Task 1 b) OLS regression of gdpgrowth on public_banks_1970
5 . // with loggdp_1960 as a control
6 .
7 . 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

------------------------------------------------------------------------------
             |               Robust
  gdpgrowth  |      Coef.   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
------------------------------------------------------------------------------

8 . // The model as a whole is not significant (F statistic) as well as the single
9 . // coefficents (t-test).
10.
11.
12. // Task 1 c) comparing the estimated coefficient of public_banks_1970 to the
13. // first row in Table V of [La Porta et al., 2002]
14.
15. // La Porta et al are reporting a negative significant coefficient (-0.0235) for
16. // GP70 while we estimate a positive and insignificant coefficient.
17. // Their result indicates that a 1% higher share of goverment ownership of banks
18. // results in average in a 2.35% lower (subsequent) per capita gdp growth.
19. // This supports the political theory.
20. // Besides the coefficent for the starting gdp level indicates possible
21. // concergence effects.
22.
23.
24.  // Task 2 a) computing summary statistics for all variables
25.
26. 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
```

```
27.
28. // comments:
29.
30. // Over all we see that the number of observations differs across the variables,
31. // this indicates missing values for some variables
32.
33. // Birth rate per 1000 population in 1970: we see that there are big differences
34. // across countries(sd)
35. // Fraction of banks owned by the Government in 1970: most imortant, there is at
36. // least one observation with a value over 100% which is not possible (max=250%)
37. // in avg the goverment of a country owns 58% of the assets of the top 10 banks
38. // in that country.
39. // Average years of schooling 1960-90: in average childs go 5years to school and
40. // the range of the observations goes from less then one year up to than 10.8
41. // Private credit to GDP in 1960: this share is in mean equal to 25%.
42. // The cross country differences are huge, what is indicated by min max and sd
43. // GDP per capita growth 1960-95: we see in average there was a small per capita
44. // in gdp growth, but again we see huge cross country differences.
45. // Log GDP per capita in 1960: here we also see that the level of the gdp per
46. // capita in 1960 differs alot across countries.
47. // Old OECD member states: we see that 28% of the countrys are old OECD member
48. // states (when we just include those with non missing values)
49.
50.
51. // Task 2 b) create scatter plot
52.
53. scatter gdpgrowth public_banks_1970

54.
55.
56. // Task 2 c) remove outlier
57.
58. // We can savely remove the county with a public_banks_1970 value of 2.5 because
59. // a goverment can not own more then 100% of the assets of the 10 greatest banks
60.
61. drop if public_banks_1970>1
  (1 observation deleted)

62.
63.
64. // Task 2 d) create scatter plot w/o outlier
65.
66. scatter gdpgrowth public_banks_1970

67.
68.
69. // Task 3 a) re run regression
70.
71. 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

------------------------------------------------------------------------------
             |               Robust
  gdpgrowth  |      Coef.   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
------------------------------------------------------------------------------
```

```
72.
73. // now after we deleted the outlier you added, we observe we have the same
74. // results as in Table V of La Porta
75.
76. // OLS is very sensetive to outliers since it is minimizing the (empirical) MSE.
77. // Observations 'far away' from 'the rest', therefore get a overproportional
78. // weight.
79.
80.
81. // Task 3 b)
82.
83. // Why are the authors using government ownership of banks in 1970 and not, for
84. // example average ownership share between 1970 and 1995?
85.
86. // One the one hand they report a relatively high correlation over time (eg .9
87. // for GP70 and GP 85) and on the other hand they argue with the availability of
88. // the data. Besides they report that the results using GP95 are very simmilar.
89.
90. // Would it be preferable to use government ownership in 1960, and if yes, why?
91.
92. // Yes, this way we can be sure that the goverment ownership is not effected by
93. // the dependend variable (exclude the possibility that the goverment ownership
94. // level is high because of low gdp, banks failed goverments bought assets to
95. // save them or some other reverse relationship). This way one could say X is
96. // predetermined.
97.
98.
99. // Task 3 c)
100
101 // Yes, the coefficent for GP75 is higly significant (p=0.003) and the
102 // coefficent for loggdp_1960 is significant on the 5% level (p=0.044.
103
104
105 // Task 3 d)
106
107 // looking on the individual effects:
108 // As the share of goverment ownership of banks increases, in average the gdp
109 // growth per capita goes down. A 1% higher goverment ownersip in 1970 results
110 // in a 2.484% lower gdp growth by capita between 1960-1995.
111 // The negatice coefficent for the gdp starting lecel indicates possible
112 // convegence effects.
113 // A 1% higher starting gdp per capita level results in avg a 0.645% lower gdp
114 // per capita growth from 1960-1995.
115
116
117 // Task 3 e)
118
119 // Our X Variables explain 12.4% of the varince of the depended variable
120 // (see R²) and the model as a whole is significant (see F statistic).
121 // Assuming that the true coefficents are all equal to zero, observing this F
122 // value or a more extreme one has a probability of less then 1%).
123
124
125 // Task 4 a)
126
127 // calc correlation matrix
128 corr
  (country_name ignored because string variable)
  (obs=80)
```

	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

```
129
130 // We have a potential problem with OVB since we see a non zero cerrelation of
131 // the two variables (we are going to include) and GP70 and simultainosly most
132 // certainly there is some relation between the two variables and our dependent
133 // variable.
134
135 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**

```
136
137
138 // Task 4 b)
139
140 // All variables are higly significant exept for schooling.
141
142
143 // Task 4 c) interpretainon
144
145 // A higher share of goverment ownership of banks results in lower gdp per
146 // capita growth in the future. That supports the political theory.
147 // A higher starting point of gdp per capita results in lower subsequent growth,
148 // like it is predicted by the theory of convergence effects.
149 // A higher birthrate also causes a loger gdp per capita growth. This supports
150 // that in the considered time window the negative effects of a increasing
151 // denumerator of gdp per capita and the temporary lower labor force due to
152 // taking care for the children outwights the effect of more furute labor force.
153
154
155 // Task 4 d)
156
157 // Test for the joint significance of schooling and birth_rate_1970.
158 // Provide H0, HA, the test statistic, its distribution and the result of the
159 // test.
160
161 // H0: both variables have a true coefficent equal to zero
162 // HA: at least one coefficent is ne 0
163 // We use a F test, the F-Statistic is follows a F distribution
164 // (when all assumptions hold)
165 // Basicly we compare the R² of a regression with and without the two variables.
166
167 test schooling birth_rate_1970
```

```
 ( 1)  schooling = 0
 ( 2)  birth_rate_1970 = 0

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

```
168
169 // The result is that we reject the H0 at the 1% lecel and that is it very
170 // unlikely that the true coefficents simultaneously equal to zero.
171
172
173 // Task 4 e)
174
175 // One option is to compare the R² for both regressions, but this we already did
176 // in the previous subtask by doing the F-Test.
177 // If we want to take account for the increase in dimension we could look ont
178 // the adj R²
179
180
181 // Task 5 a)
182
183
184 // Goverment owned banks are maybe not efficent when they are the only lender in
185 // the financial system, but they are maybe a efficent comlement to a well
186 // working financal market in the sense that they can step in when market
187 // failtures occur.
188 // The private credit to gdp can be used as a mesurement for the development of
189 // the financial market.
190
191
192 // Task 5 b)
193
194 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

```
195
196
197 // Task 5 c)
198
199 // When a country has zero private credit, then a 1% higher share of goverment
200 // ownership of banks results in avg in a 3.97% decrease in the subsequent gdp
201 // per capita growth.
202 // when a country has zero goverment ownership of banks, then a 1% higher
203 // privatecredit of gdp share results in avg in a 1.86% increase in the
204 // subsequent gdp per capita growth.
```

```
205
206
207 // Task 5 d)
208
209
210 // here we can only give a qualitative interpretaion or make pointwise
211 // comparisons, meaning compare the Prediction for different X varaible values.
212 // The higher the share of private credit to gdp, the better the marginal effect
213 // of public banks. The marginal effect is even positive given a value
214 // of private credit higher than 58% ( but this isn't significant at that point)
215
216
217 // Task 5 e)
218
219 ge margin = -.0397587+.254575*.0685466

220 drop margin

221 //or
222 margins, dydx(public_banks_1970) atmeans

  Conditional marginal effects                    Number of obs     =         82
  Model VCE    : Robust

  Expression   : Linear prediction, predict()
  dy/dx w.r.t. : public_banks_1970
  at           : public_~1970    =    .5556605 (mean)
                 private~1960    =     .254575 (mean)
                 loggdp_1960     =    5.849251 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
public_~1970 |  -.0223084    .007252    -3.08   0.003    -.0367489   -.0078679
------------------------------------------------------------------------------

223 // -.0223085 is the marginal effect evaluatet at the mean of private credit
224 // (and all other vars).
225
226
227 // Task 5 f)
228
229 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))

  Average marginal effects                        Number of obs     =         82
  Model VCE    : Robust

  Expression   : Linear prediction, predict()
  dy/dx w.r.t. : public_banks_1970

  1._at        : private~1960    =          .1

  2._at        : private~1960    =          .2

  3._at        : private~1960    =          .3

  4._at        : private~1960    =          .4

  5._at        : private~1960    =          .5

  6._at        : private~1960    =          .6

  7._at        : private~1960    =          .7

  8._at        : private~1960    =          .8

  9._at        : private~1960    =          .9

  10._at       : private~1960    =           1
```

```
11._at          : private~1960    =         1.1

12._at          : private~1960    =         1.2
```

	Delta-method dy/dx	Std. Err.	t	P>|t|	[95% Conf.	Interval]
public_~1970						
_at						
1	-.032904	.009173	-3.59	0.001	-.0511697	-.0146383
2	-.0260493	.0078525	-3.32	0.001	-.0416857	-.010413
3	-.0191947	.0068384	-2.81	0.006	-.0328117	-.0055777
4	-.01234	.0062808	-1.96	0.053	-.0248468	.0001667
5	-.0054854	.0063022	-0.87	0.387	-.0180347	.0070639
6	.0013693	.0068971	0.20	0.843	-.0123646	.0151032
7	.0082239	.0079376	1.04	0.303	-.0075819	.0240298
8	.0150786	.009275	1.63	0.108	-.0033903	.0335475
9	.0219333	.0107995	2.03	0.046	.0004288	.0434377
10	.0287879	.0124425	2.31	0.023	.0040118	.053564
11	.0356426	.0141628	2.52	0.014	.0074409	.0638443
12	.0424972	.0159354	2.67	0.009	.0107657	.0742287

```
230 marginsplot, level(95)

    Variables that uniquely identify margins: private_credit_1960

231
232 // the marginal effect of GP70 isn't significant in the range of approxximately
233 // 40% to 90% for private credit in 1960.
234
235
236 // Task 6 a)
237
238 // oecd or noecd will be exluded due to multicolinearity we, out of the set
239 // intercept oecd and noecd always one variabel can be expressed as a linear
240 // combination of the other two.
241
242 ge noecd = (oecd==0)

243 reg gdpgrowth noecd i.oecd##c.public_banks_1970 i.oecd##c.loggdp_1960
  note: 1.oecd omitted because of collinearity
```

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

```
244
245
246 // Task 6 b)
247
248 // since gdp level in 1960 >0 it makes no sense to look on the constant
249 // seperately.
250 // The coefficent for noecd dummy can not be interpreted seperately since the
251 // normal mean comparison makes no sense when we do not take account for our
252 // interaction terms. We should instead look on marginal effects
253
254
255 // Task 6 c)
256
257 // a clean interpretation of the interaction coefficents is also only pointwise
258 // possible.
259 // Due to similar reasons as in b)
260
261
262 // Task 7 a)
263
264 reg gdpgrowth noecd i.oecd##c.public_banks_1970 i.oecd##c.loggdp_1960, robust nocons
  > tant
```

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**

```
265
266
267 // Task 7 b)
268
269
270 // the coefficent for the dummy OECD is the intercept for the oecd observations
271 // and the coefficent for the dummy no OECD is the intercept for the no oecd
272 // countries.
273 // The coefficent of the intersection of public banks and oecd gives us the the
```

```
274 // difference in the effect of a marginal increase in public banks between oecd
275 // and no oecd countries a one percent increase in the share of goverment owned
276 // banks increases the per capita gdp growth by 2.64% more when the
277 // country is a old oecd member state compared to a non member state.
278 // The coefficent of the intersection of loggdp_1960 and oecd gives us the the
279 // difference in the effect of a marginal increase in loggdp_1960 between oecd
280 // and no oecd countries, this coefficent is not significant (alone).
281
282 // to summarize we got the same results as when we run two regressions, one for
283 // oecd members and one for non oecd members.
284 // But we put the results together in one model - the fitted values are exactly
285 // the same for our model or the two seperated regressions.
286
287
288 // Task 7 c)
289
290 // We have two dummies where always exactly one of them is equal to one, so if
291 // we include a constant we would get trouble with colinearity.
292 // The constant would be a linear combination of the two dummies.
293
294
295 // Task 7 d)
296
297 // In the previous regression we had the problem with to colinearity of the
298 // dummies and the constant, which stata solved by omitting one dummy.
299 // But the version of the model with a dummy and a intercept(Regression 5) is
300 // not that clear to interpretate than the solution with two dmmies(Regression6)
301
302
303 // closing log-File
304 log close
       name:  <unnamed>
        log:  C:\Users\s3sibong\Downloads\Tutorial2.smcl
   log type:  smcl
  closed on:   2 May 2020, 13:58:57
-------------------------------------------------------------------------------
```