# Tutorial 2
# Specification and Interaction Terms - Due April 26th, 2022 20:00h

Empirical Banking and Finance
Konrad Adler

Summer, 2022

This excercise is based[1] on the articles by [La Porta et al., 2002] and [Koerner and Schnabel, 2011]. The original data can be found on Rafael La Porta's Webpage.

1. Regression 1
   a) Please load the file *dataEmpBF_Tutorial2.dta* into STATA.
   b) Run an OLS regression of *gdpgrowth* on *public_banks_1970* with *loggdp_1960* as a control.
   c) Compare the estimated coefficient of *public_banks_1970* to the first column in Table V of [La Porta et al., 2002]. Comment very briefly.
2. Descriptive statistics
   a) Compute summary statistics for all the variables in the dataset. Comment briefly your findings.
   b) Create a scatterplot of *gdpgrowth* and *public_banks_1970*.
   c) Exclude the (2) outliers. Why can we safely assume that there is a problem with these two observations? From now on work with the dataset without the outliers.
   d) Re-create the scatterplot.
3. Regression 2
   a) Re-run OLS regression 1. Compare your results again to the first column in Table V of [La Porta et al., 2002]. Comment and explain the intuition in two sentences why OLS is so sensitive to outliers.
   b) Are the coefficients statistically significant?
   c) Interpret the sign of each coefficient.
   d) How good are the $X$ variables at explaining *gdpgrowth*? Which measure is usually considered?

[1] Thanks to Ulrich Schüwer

e) Why could it be problematic to use government ownership measured in 1970?

4. Regression 3: Control variables

   a) Add the variables *schooling* and *birth_rate_1970* to Regression 2. Why does it make sense to include these variables as control variables? *Hint: think about and use the two-step procedure shown in class*

   b) Which coefficients are statistically significant?

   c) Interpret the sign of the coefficients if they are statistically significant. What is the economic size implied by these coefficients?

   d) Test for the joint significance of *schooling* and *birth_rate_1970*. Provide $H_0$, $H_A$, the test statistic, its distribution and the result of the test.

   e) Compare the goodness of fit to regression 2. Which measure do you use?

5. Regression 4: Interactions

   a) [Koerner and Schnabel, 2011] suggest interacting public ownership of banks with the initial level of private credit to GDP. Without running the regression yet, provide arguments why this might be a good idea.

   b) Run the regression of *gdpgrowth* on *public_banks_1970*, *private_credit_1960*, the interaction between the latter two and *loggdp_1960*.

   c) Interpret the sign of *public_banks_1970* and *private_credit_1960*. What exactly do they measure?

   d) Interpret the sign of the interaction term.

   e) What is the impact of an increase in public ownership of banks on gdp growth holding all other variables fixed at their mean ? *Hint: Use the margins command*

   f) How does the impact of *public_banks_1970* on *gdpgrowth* vary over different levels of finanical development (as measured by *private_credit_1960*)? Where is the effect statistically significant? *Hint: Use the margins and the marginsplot command*

6. Regression 5: Dummies I

   a) Run a new regression of *gdpgrowth* on *public_banks_1970* and *loggdp_1960*. Add the *europe* dummy variable and interactions of both $X$ variables with the *europe* dummy variable.

   b) What is the interpetation of the constant? What is the interpretation of the *europe* dummy?

   c) What is the interpretation of the interaction term coefficients?

   d) Someone suggests that you don't have to include the *europe* dummy in the previous regression. Re-run the previous regression without the *europe* dummy and provide an answer to the suggestion.

7. Regression 6: Dummies II

a) Run a new regression of *gdpgrowth* on *public_banks_1970* and *loggdp_1960*. Interact both LHS variables once with the *europe* dummy and once with a non-European country dummy. Also include both dummies, but exclude the constant.

b) How do you interpret the two dummies?

c) Why do you have to exclude the constant? What happens if you keep it?

d) Run two separate regressions of *gdpgrowth* on *public_banks_1970* and *loggdp_1960*. Once only for the *europe*=1 sample and once for all other countries. Compare the coefficients to the results in a).

e) How are the results different from Regression 5?

## References

[Koerner and Schnabel, 2011] Koerner, T. and Schnabel, I. (2011). Public ownership of banks and economic growth. *The Economics of Transition*, 19(3):407–441.

[La Porta et al., 2002] La Porta, R., Lopez-De-Silanes, F., and Shleifer, A. (2002). Government ownership of banks. *The Journal of Finance*, 57(1):265–301.