Statistical analyses, theory and practice in social sciences: an example of complex relations in the social reality

Quantitative Methods

When we think of democratic, rich countries, we often think of 'good countries': stable countries without (violent) conflicts. Not just in common sense, but also by looking at correlation coeffiecients we can say that these two features are highly related. However, can we state that these factors are statistically significant related? In other words, can we say that a high GDP in a democratic regime leads to a more politically stable situation in terms of the absence of violent conflicts? It is expected that there is a positive relation between GDP and political stability, whereby the hypotheses are:

H0: A higher GDP does not lead to a more stable political situation in democracies.
H1: A higher GDP does lead to a more stable political situation in democracies.

Research has been done on this relationship before, so it is not a new question we are asking (see for example Bingham Powell, 1982; Feng, 1997). However, the aim of this essay is to show how regression analysis can be used in the field of political science and how it can contribute to the 'real world' around us: the knowledge is valuable to practitioners, organizations and citizens because it helps them understand the world around them, and maybe even enables them to change it. Therefore, the formulated hypotheses are relevant in scientific and in societal terms.

Data & method

The Pippa Norris *Democracy Cross-national Data* (release 3.0, spring 2009) dataset consists of 1002 variables, with 109 observations from countries around the world. Norris collected

the data for this dataset herself, with original resources as the World Values Survey, the World Bank, the United Nations and Freedom House. The characters of the variables range from pure informative variables describing features of the countries to political economic variables, giving information on, for example, the level of democracy and the political structure of the country. A selection is made in countries which are labeled *free* by the Freedom House democracy index in order to correct for politically stable, but authoritarian, regimes.

The two main variables in this analysis are: political stability and GDP. To give a valid and reliable answer to the research question, a third potential explanatory variable is included in the model. This control variable is 'voice and accountability'. The dataset used for this analysis has a selection of countries which are indicated *free* in the Freedom House index. However, we still have to control for a false effect on political stability in countries where citizens do not have a lot of opportunities to express their opinions, rights to demonstrate and where media is restricted. In other words, we have to control for a false effect on political stability in democratic regimes where the level of 'voice and accountability' is relatively low (Kaufmann, 2010. See also Mishler & Rose, 2001). Moreover, we have to include this third variable as a control variable because the Freedom House index is not uncontroversial in the field of political science (see for example Armstrong, 2011).

The method used in this analysis is multiple regression. This way, the effect of variation in the independent variable on the dependent variable is explored while taking into account the explanatory power of another independent variable: the control variable

(Agresti & Franklin, 2009: 123). Regression analysis comes with assumptions, and in order to run a valid analysis it is important to test for characteristics in the data and of the variables: linearity, homoscedasticity, normality and independence of observations (errors) (Meier, Brudney and Bohte, 2006). The results of the tests[1] show that the data and variables fit the assumptions and characteristics needed for a valid regression analysis.

Analysis

When looking at the correlation coefficient (0.403) between the two variables, we can state that GDP and political stability are highly correlated. To look at the relationship and explanatory power of the different variables however, we have to use regression models. The two regression models are displayed in table 1.

In the first model, the two main variables are included: political stability and GDP. The second model is a complete model with also the control variable of voice and accountability. As we can see, the relation between GDP and political stability is a complex one. In the first model, which explains 16% of the variance in political stability in the democracies, GDP has a statistically significant effect, but the coefficient is zero; which means that we cannot say anything about the *actual effect* of an increasing GDP on political stability.

[1] The commands for testing for the regression assumptions are included in the STATA *do-file*.
Besides the four assumptions for regression (linearity, homoscedasticity, normality and independence of observations/erros) there is also a test included on the multicollinearity in the dataset because we run a multiple regression analysis. With this test, variance inflation factor (VIF), we can control whether or not the different independent variables explain the same variance of the dependent variable. This would be problematic because false conclusions would be made on the relationship and strength of it if that is the case.

Table 1: Effects of GDP on level of political stability

	Model 1		Model 2	
Intercept	0.377	***	-0.127	
	(0.085)		(0.097)	
GDP	0.000	***	0.0~7	
	(∞)		(∞)	
Voice and Accountability			0.982	***
			(0.136)	
R^2	*0.16*		*0.48*	
N	89		89	

*Note: Effects are unstandardized coefficients. The statistical significance is presented as: *** p<0.001; ** p<0.01; *p<0.05. Standard errors are displayed between brackets.*

The second model, where the variable 'voice and accountability' is also included, does not provide us with more information about the effect of GDP on political stability either. It does provide us with information about the relation between voice and accountability and political stability. Voice and accountability has a statistically significant positive effect on political stability with a coefficient of 0.982. This means that when a democratic regime scores one point higher on the Voice and Accountability index of Kaufmann, the political stability in a country increases with 0.982 - almost 1 point. That means that we can state, with a level of statistical significance of 0.05, that a higher level of voice and accountability

leads to a higher level political stability. This coincides with the observation made in the introduction, the freeer, richer and (politically) developed a country is, how more likely it is we are talking about a stable country.

Conclusion and interpretation

The analysis showed how one can use statistical analysis and tests in social science, and how to interpret results of a regression analysis have to be interpreted. We cannot reject the hypothesis that a higher GDP leads to more political stability in democratic regimes because the results of the analysis is statistically significant, although we do not know how strong this relationship is.

When we included a third variable, voice and accountability, there was no significant effect of GDP on political stability anymore. This means that, when a democratic regime has a high level of 'voice and accountability', there is a high chance that the country has a high level of political stability too - where the level of GDP does not have a statistically significant effect anymore. Theoretically it is hard to defend this positive relationship - this reflects the pitfalls and shortcomings of this method of analysis.

What we can conclude is that both GDP and the level of 'voice and accountability' refer to a specific form of 'developed' ('modern') countries with democratic regimes: developed countries with high political stability. And this makes us return to the first statement of this essay: When we think of democratic, rich countries, we often think of 'good countries': stable countries without (violent) conflicts. What the determinates of this are is still difficult to say and will always be: it is a classic example of a complex set of interrelated features in social sciences.

References

Agresti & Franklin (2009) *Statistics: the art and science of learning from data*. Pearson education.

Armstrong, D.A. (2011) Stability and change in the Freedom House political rights and civil liberties measures. *Journal of Peace Research* 48(5): 653-662.

Bingham Powell, G. (jr.) (1982) *Contemporary Democracies: Participation, Stability and Violence*. Cambridge: Harvard College.

Feng, Y. (1997) Democracy, Political Stability and Economic Growth. *British Journal of Political Science* 27(3): 391 - 418.

Meier, K.J., Brudney, J.L., Bohte, J. (2006) *Applied Statistics for Public and Non-Profit Administration*. Belmonth, California: Thomson/Wadsworth.

Mishler, W. & Rose, R. (2001) Political Support for Incomplete Democracies: Realist vs. Idealist Theories and Measures. *International Political Science Review* 22: 303.

Kaufmann, D. (2010) Governance Matters 2010: Worldwide Governance Indicators Highlight Governance Successes, Reversals, and Failures. http://www.brookings.edu/research/opinions/2010/09/24-wgi-kaufmann. Last accessed on 22 November 2013.

Norris, P. (2009) *Democracy Cross-national Data* release 3. STATA file, download from http://www.hks.harvard.edu/fs/pnorris/Data/Data.htm. Last accessed on 22 November 2013.

APPENDIX: STATA *do-file*

```
use "C:\Users\Administrator\Downloads\Democracy Crossnational Data Spring 2009
>  StataSE (4).dta"
describe TypeDemo2007
summarize TypeDemo2007
tab TypeDemo2007
drop if TypeDemo2007 == 2
drop if TypeDemo2007 == 3
describe Stable2006
summarize Stable2006
tab Stable2006
describe GDP2006
summarize GDP2006
tab GDP2006
describe Voice2006
summarize Voice2006
tab Voice2006
corr Stable2006 GDP2006
graph matrix GDP2006 Stable2006 Voice2006
regress Stable2006 GDP2006
predict r, resid
kdensity r, normal
estat imtest
rvfplot
vif
estimates store model1
regress Stable2006 GDP2006 Voice2006
predict r2, resid
```

```
kdensity r2, normal
estat imtest
rvfplot
vif
estimates store model2
```