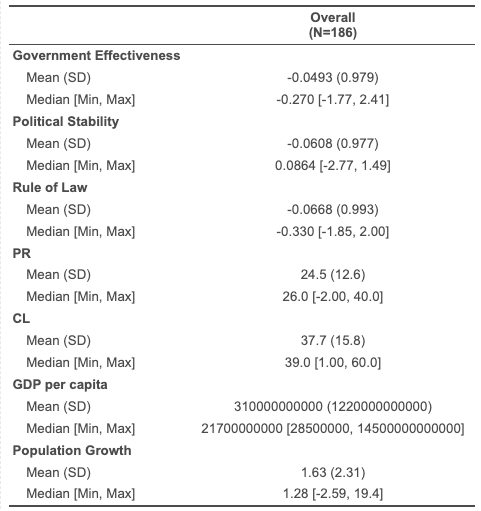
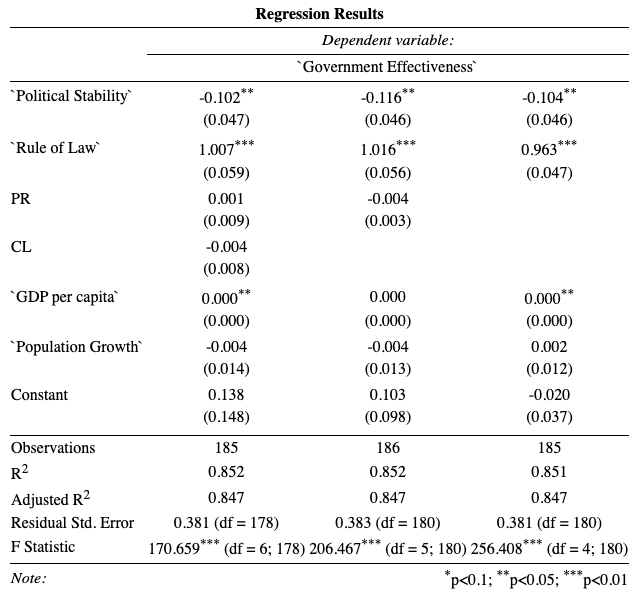
***Pavlenko Mariia,***

***BPS-222.***

# Block 2. Table with descriptive statistics



# Block 3. Table with the results of the OLS regressions and its interpretation



1. **Are there any statistically significant independent variables in the model of the best fit?**

Model3 is the best in terms of prediction because it has the lowest AIC, AICc and BIC. Model1 and Model2 are a bit better in terms of explanation because their R2 is 0.002 higher than Model3's. All models have the same R2 adjusted, so we cannot determine the best one in this regard. Model1 and Model2 are also the tiny bit better in terms of RMSE as theirs is 0.002 smaller than Model3's.

🡪 However, I think that the difference in 0.002 is not that significant, considering Model3's prediction ability, so Model3 could be rightfully called the best fit one out of three.

There are significant independent variables in the model of the best fit: *Political Stability* and *Rule of Law.* The first one is significant on a 5% confidence level and the second one on 99.9% confidence level.

1. **What is the direction of relationship between each of the independent variables and dependent variable and what does the estimate for each of the IVs essentially mean?**

The direction of relationship between each of independent variables and dependent variable:

1. *Government Effectiveness ~ Political Stability*: they have a negative relationship with the estimate of -2.093e-03 which means that *Government Effectiveness* changes by -2.093e-03 with a one-unit change in *Political stability*.
2. *Government Effectiveness ~ Rule of Law*: they have a positive relationship with an estimate of 9.803e-01 which means that *Government Effectiveness* changes by 9.803e-01 with a one-unit change in *Rule of Law*
3. *Government Effectiveness ~ GDP per capita*: they have a positive relationship with an estimate of 3.143e-14 which means that *Government Effectiveness* changes by 3.143e-14 with a one-unit change in *GDP per capita*
4. *Government Effectiveness ~ Population Growth*: they have a positive relationship with an estimate of 1.015e-03 which means that *Government Effectiveness* changes by 1.015e-03 with a one-unit change in *Population Growth*
5. **What is the predictive capability of the model of the best fit?**

According to Adjusted R-squared the model’s predictive capability is 0.8471 which is pretty high and means that the model is able to generalize well and explains over 80% of variation in the dependent variable.

1. **What does F-statistic say about the model’s characteristics?**

F-statistic indicates whether there is a significant relationship between the independent variables and the dependent variable in the model. It also shows how well the independent variables collectively explain the variation in the dependent variable against a scenario where there is no relationship between the independent and dependent variables.

In our case, the F-statistic is 257.3 with degrees of freedom (DF) 4 and 181 which is a pretty good score and considering a small (<0.01) p-value of < 2.2e-16 and higher than 4 degrees of freedom which means that the model is statistically significant and has a pretty high predictive capability.

**[Block 4. OLS regression diagnostics]**

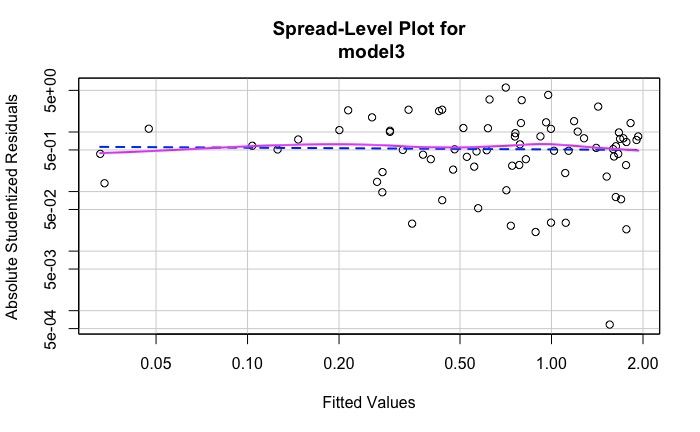
1. **Multicollinearity**

In order to determine that multicollinearity exists in the model, we should look at vif value which, in turn, should exceed 10. But in our case the value is below 10 for all of the variables evaluated. Therefore, there was no multicollinearity found in the model of the best fit.

1. **Heteroscedasticity**

On the plot both lines are shaky, we can assume the problem of non-constant DV variance aka heteroscedasticity but cannot determine that for sure just from the plot.

However, based on ncvTest which has shown a p-value of 4.312e-05 (<0.05), we can be sure that there is heteroscedasticity in the model that needs to be fixed. That was done using vcovHC function by creating a covariation matrix with robust errors.

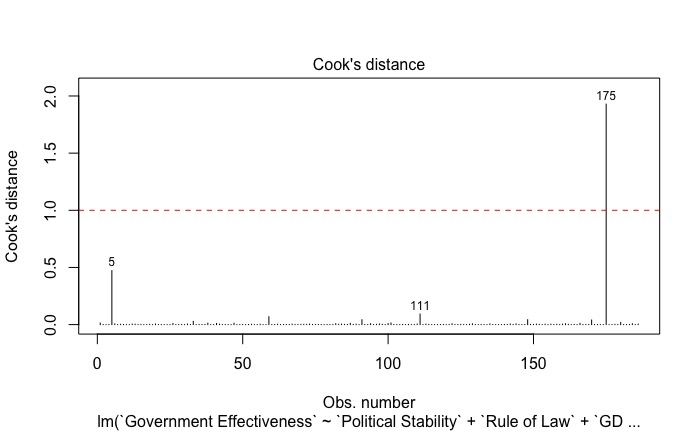
****

1. **Outliers**

outlierTest was able to detect two outliers in the model with the row numbers of 111 and 59 which had Bonferroni p-value of less than 0.05 (Marshall Islands and Federative States of Micronesia respectively)

1. **Influential observations**

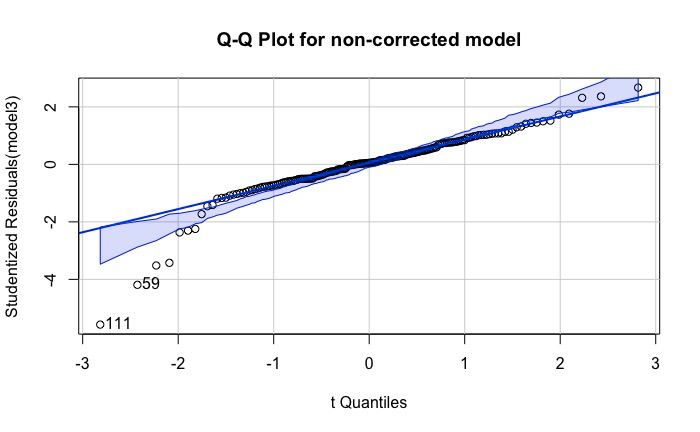
On the Cook’s distance plot the outliers do not happen to cross the default distance line which means that they are not influential observations. However, there is an observation with a row number of 175 that crosses said line and reaches a distance of 2.0 which means that it is a rather influential observation and we are deleting it from the model. Probably, the outlier test was not able to detect it for technical reasons, however, Cook’s distance uses different measurement and was able to detect it. Said observation is the United States of America, so it is not that surprising after all.



1. **Non-normality of regression residuals distribution**

There are point on the plot that are beyond the boundaries of the two lines which are confidence intervals. This means that we can assume the violation of the assumption about the normality of distribution.

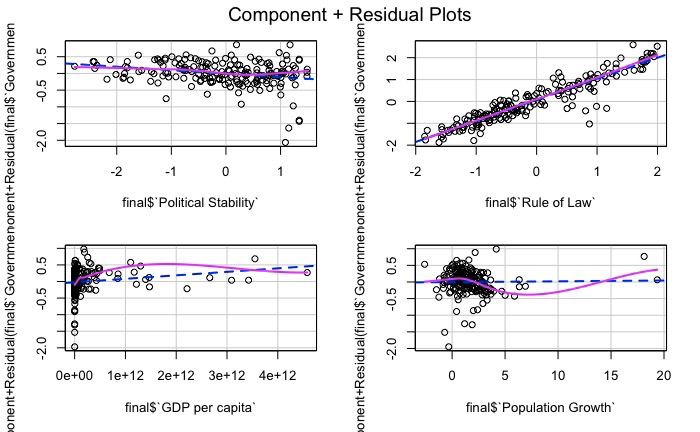
And the test (powerTransform) proves it. The p-value for lambda = (1) happen to be less than 0.05 (2.22e-16 ) which means that the distribution of regression residuals' is not normal.



1. **Non-linearity analysis**

Judging by the plot, we can assume mom-linearity for variables: Political Stability, GDP per capita and Population Growth because their trendlines do not match. The trendlines for the Rule of Law seem to be matching, however. Though, we cannot make conclusions from the plot only.

The p-value for MLE of lambda is less than 0.05 for Rule of Law and GDP per capita which means that there is non-linearity present for these variables and it needs to be fixed.



1. **Interaction effect**

Based on the comparison of the additive and interaction effect models, it turned out that additive model is still better based on AIC, BIC, AICc, R2, R2-adjusted and RMSE scores.