Chapter 2 - Political Stability Around the World Since the End of the Cold War: Descriptive Statistics and Exploratory Data Analysis

Adrien Ratsimbaharison

10/14/2018

Outline:

2.1 - Descriptive Statistics

2.2 - Exploratory Data Analysis

## 2.1 - Descriptive Statistics

**General Stability of the World Since the End of the Cold War**

Following Kaufman et al.’s methodology (see World Bank 2016; and Kaufmann, Kraay, and Mastruzzi 2011), the political stability estimate of a given country for a given year is based on the aggregation of the perceptions (or assessments) from several survey respondents, including the following so-called five representative sources: Economist Intelligence Unit Riskwire & Democracy Index (EIU), World Economic Forum Global Competitiveness Report (GCS), Cingranelli Richards Human Rights Database and Political Terror Scale (HUM), Institutional Profiles Database (IPD), Political Risk Services International Country Risk Guide (PRS), and Global Insight Business Conditions and Risk Indicators (WMO). The annual political stability estimate ranges from approximately -2.5 to 2.5, with some extreme values less than -2.5. Using the dataset on the political stability estimate of all countries in the world for the period of 1996 to 2016, we can compute the stability of the world for the entire period and the annual stability trend from 1996 to 2016. Furthemore, we can compare the stability of the different regions in the world.

Thus, based on the dataset on the political stability estimate of all countries in the world, we can state that the whole world was not stable since the end of the Cold War. Indeed, as shown in Table 1 and Figure 1 below, the summary statistics of political stability during the time period of 1996-2016 around the world produces an average of -0.0199 with standard deviation of 1.00, which seems to indicate that the whole world has been slightly unstable during this time period. One can also notice that whereas the maximum value of political stability is 1.96, the minimum value is -3.31.

Furthermore, Figure 2 shows that the data on political stability is not normally distributed, but skweed to the left, which indicates the existence of outliers with negative values. Figure 2 also shows that the mode is between 0.5 and 1.

# a clean dataset including WGI, development indicators, and RegimeType indicators, is available in the local repository under "WGIdevRegimeType.csv"  
  
  
# Table 1 - Summary statistics of the political stability estimate  
  
summary(WGIdevRegimeType$stability, na.rm = TRUE)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -2.84465 -0.80456 -0.09958 -0.17179 0.55980 1.76010

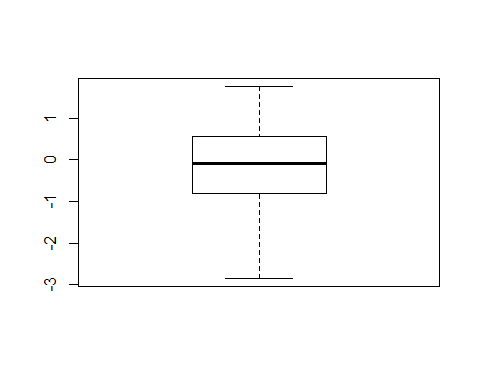


Fig. 1 - Political Stability: Summary Statistics

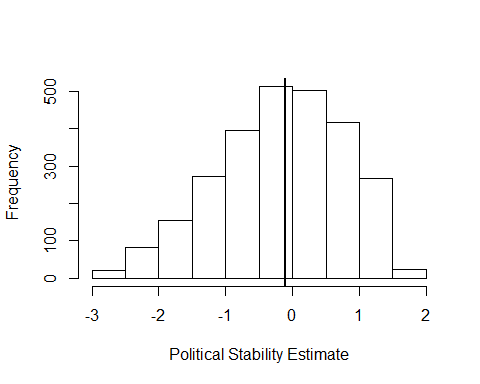


Figure 2 - Political Stability: Frequency

**Political Stability Trend: Annual Average Around the World**

As shown in Table 2 and Figure 3, the annual average of politial stability has remained below zero during the time period of 1996-2016 throughout the world. There was no time it had a positive value.

Table 2 - Political Stability Trend: Annual Average Around the World

|  |  |
| --- | --- |
| date | AnnualAverage |
| 1996 | -0.0398903 |
| 1998 | -0.0424399 |
| 2000 | -0.0978401 |
| 2002 | -0.0883550 |
| 2003 | -0.1512978 |
| 2004 | -0.2021411 |
| 2005 | -0.1981306 |
| 2006 | -0.1892264 |
| 2007 | -0.1801231 |
| 2008 | -0.1858479 |
| 2009 | -0.1969126 |
| 2010 | -0.2115577 |
| 2011 | -0.2122642 |
| 2012 | -0.1981355 |
| 2013 | -0.1992925 |
| 2014 | -0.2143603 |
| 2015 | -0.2122176 |
| 2016 | -0.2192850 |

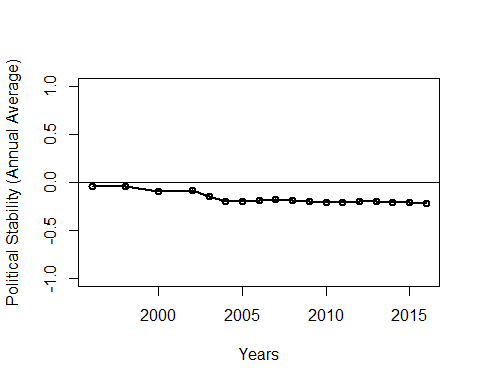


Fig. 3 - Political Stability Trend: Annual Average Around the World

**Political Stability: Average by Region**

Table 3 and Figure 4 clearly show that the regions of Africa and Asia, with respectively an average of -0.575 and -0.388, were the two most unstable regions in the world during the time period of 1996-2016. On the contrary, the regions of Oceania and Europe, with respectively an average of 0.820 and 0.651, were the two most unstable regions in the world

Table 3 - Political Stability: Average by Region

|  |  |  |
| --- | --- | --- |
| region | mean | sd |
| Africa | -0.5904448 | 0.8785006 |
| Americas | -0.1512632 | 0.6584877 |
| Asia | -0.4587124 | 0.9467938 |
| Europe | 0.6308342 | 0.6313327 |
| Oceania | 0.4445214 | 0.7353292 |

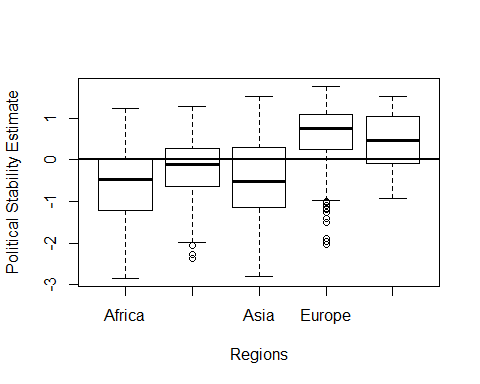


Fig. 4 - Political Stability: Average by Region

## 2.2 - Exploratory Data Analysis

*Correlation Between Political Stability and the other World Governance Indicators (WGI)*

In this section, we plot political stability against the other world governance indicators (corruption control, government effectiveness, regulatory quality, rule of law, and voice and accountability), and find that political stability is strongly and significantly correlated with these other world governance indicators. The correlation coefficients range from 0.63 to 0.77, with a p-value close to zero.

This result allows us to include all world governance indicators in our prediction models.

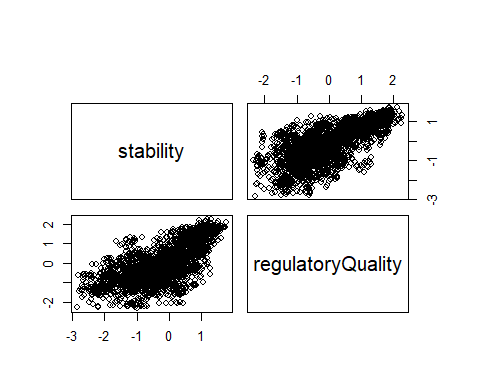


Fig. 5 - Correlation between Political Stability and the other World Governance Indicators (WGI)

#Table 4 - Correlation Matrix between Political Stability and the other World Governance Indicators (WGI)  
  
correlationWGI

## stability regulatoryQuality  
## stability 1.00 0.69  
## regulatoryQuality 0.69 1.00  
##   
## n= 2647   
##   
##   
## P  
## stability regulatoryQuality  
## stability 0   
## regulatoryQuality 0

### *Correlation Between Political Stability, Development, and Social Inequality*

In this section, we plot political stability against the indicators of socioeconomic development (GNI per capita and Human Development Index (HDI)) and social inequality (GINI index), and find that while the correlation between political stability and GNI per capita is relatively strong (0.61), those between political stability, on the one hand, and HID and GNI, on the other hand, are moderate (respectively 0.48 and -0.34).

Nevertheless, since all of these correlations are statistically significant (with p-values close to zero), we can still include these indicators of socioeconomic developmentin and social inequality in our prediction models.

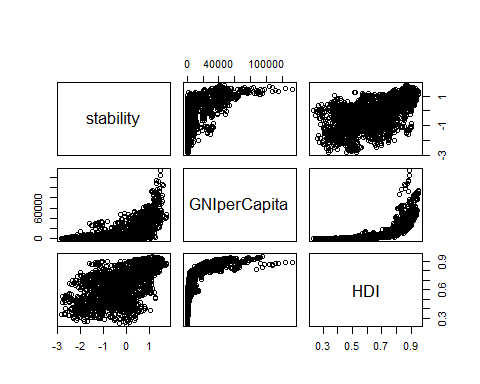


Fig. 6 - Correlation between Political Stability, Development, and Social Inequality

#Table 5 - Correlation Matrix between Political Stability and the Indicators of Development and Social Inequality  
  
correlation.devStability

## stability GNIperCapita HDI  
## stability 1.00 0.6 0.61  
## GNIperCapita 0.60 1.0 0.70  
## HDI 0.61 0.7 1.00  
##   
## n= 2647   
##   
##   
## P  
## stability GNIperCapita HDI  
## stability 0 0   
## GNIperCapita 0 0   
## HDI 0 0

### *Correlation Between Political Instability and Regime Type Indicators*

In this section, we plot political stability against the indicators of regime type (Polity IV revised combined polity score, Polity IV score on institutionalized democracy, Polity IV score on institutionalized autocracy, Freedom House inversed score on political rights, Freedom House inversed score on civil liberties, Freedom House combined score), and find that while the correlation between political stability and GNI per capita is relatively strong (0.61), those between political stability, on the one hand, and HID and GNI, on the other hand, are moderate (respectively 0.48 and -0.34).

Nevertheless, since all of these correlations are statistically significant (with p-values close to zero), we can still include these indicators of socioeconomic developmentin and social inequality in our prediction models.

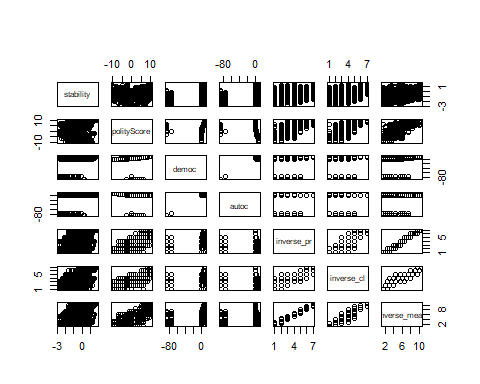


Fig. 7 - Political Instability and Regime Type Indicators

#Table 6 - Correlation Matrix between Political Stability and Regime Type Indicators  
  
Correlation.regimeTypeStability

## stability polityScore democ autoc inverse\_pr inverse\_cl  
## stability 1.00 0.34 0.35 0.20 0.55 0.62  
## polityScore 0.34 1.00 0.33 -0.12 0.88 0.84  
## democ 0.35 0.33 1.00 0.89 0.39 0.39  
## autoc 0.20 -0.12 0.89 1.00 -0.01 0.01  
## inverse\_pr 0.55 0.88 0.39 -0.01 1.00 0.94  
## inverse\_cl 0.62 0.84 0.39 0.01 0.94 1.00  
## inverse\_mean 0.58 0.88 0.39 -0.01 0.99 0.97  
## inverse\_mean  
## stability 0.58  
## polityScore 0.88  
## democ 0.39  
## autoc -0.01  
## inverse\_pr 0.99  
## inverse\_cl 0.97  
## inverse\_mean 1.00  
##   
## n= 2647   
##   
##   
## P  
## stability polityScore democ autoc inverse\_pr inverse\_cl  
## stability 0.0000 0.0000 0.0000 0.0000 0.0000   
## polityScore 0.0000 0.0000 0.0000 0.0000 0.0000   
## democ 0.0000 0.0000 0.0000 0.0000 0.0000   
## autoc 0.0000 0.0000 0.0000 0.5551 0.6360   
## inverse\_pr 0.0000 0.0000 0.0000 0.5551 0.0000   
## inverse\_cl 0.0000 0.0000 0.0000 0.6360 0.0000   
## inverse\_mean 0.0000 0.0000 0.0000 0.7799 0.0000 0.0000   
## inverse\_mean  
## stability 0.0000   
## polityScore 0.0000   
## democ 0.0000   
## autoc 0.7799   
## inverse\_pr 0.0000   
## inverse\_cl 0.0000   
## inverse\_mean

### *Generalized Linear Regression Model to Identify the Significant Variables*

In this section, we include all variables in a single dataframe and run a step function in order to identify the most significant variables to be included in the prediction model.

However, before runing the step function, we need to clean the data. The following steps were followed in order to remove the missing values (NAs):

* first, we identify the variables with high percentage of missing values (more than 20%) and remove them from the dataframe,
* second, from the remaining variables and observations, we remove the incomplete cases and only use the complete ones.

# Table 7 - Summary of the Generalized Linear Regression Model Using Step Function  
# modelFit1 <- step(glm( data=WGIdevRegimeType, stability ~ .), trace=0,steps=10000)   
# summary(modelFit1)

The above step function allows us the identify the following variables as the most significant predictors to be included in the prediction models: subregion, regulatoryQuality, GNIperCapita, civilLiberties, FHstatus.

## III. Predicting Political Stability Using Machine Learning

### *Predicting with Generalized Linear Regression*

Using the variables identified through the above step function, we run machine learning predictions using generalized linear regression with the caret package, and we obtain the following results:

* As shown in Table 8a, on the trainning data set, the following variables are still significant: subregion, regulatoryQuality, GNIperCapita, civilLiberties, FHstatus.
* As shown in Table 8b, on the test data set, the same variables (subregion, regulatoryQuality, GNIperCapita, civilLiberties, FHstatus) are also significant.

#Table 8a - Machine Learning Prediction Using Generalized Linear Regression on the Trainning Dataset  
summary(modelFit1)

##   
## Call:  
## glm(formula = stability ~ ., data = train1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.40432 -0.32983 0.03868 0.38201 2.16063   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.130e+00 1.153e-01 -18.476 < 2e-16 \*\*\*  
## regulatoryQuality 1.768e-01 3.002e-02 5.888 4.59e-09 \*\*\*  
## GNIperCapita 5.620e-06 1.156e-06 4.860 1.26e-06 \*\*\*  
## HDI 4.664e-01 1.330e-01 3.508 0.000462 \*\*\*  
## polityScore 5.431e-02 4.463e-02 1.217 0.223847   
## democ -1.972e-02 4.648e-02 -0.424 0.671457   
## autoc 2.070e-01 4.588e-02 4.512 6.80e-06 \*\*\*  
## xrreg -9.861e-02 3.611e-02 -2.731 0.006372 \*\*   
## xrcomp -8.158e-02 3.663e-02 -2.227 0.026057 \*   
## inverse\_pr 4.516e-02 2.147e-02 2.103 0.035563 \*   
## inverse\_cl 2.982e-01 2.433e-02 12.258 < 2e-16 \*\*\*  
## inverse\_mean NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.355487)  
##   
## Null deviance: 1742.48 on 1986 degrees of freedom  
## Residual deviance: 702.44 on 1976 degrees of freedom  
## AIC: 3596.7  
##   
## Number of Fisher Scoring iterations: 2

# Table 8b - Machine Learning Prediction Using Generalized Linear Regression on the Testing Dataset  
summary(modelFit2)

##   
## Call:  
## glm(formula = stability ~ ., data = test1)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.05814 -0.38413 0.02912 0.41472 2.12327   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.158e+00 2.112e-01 -10.221 < 2e-16 \*\*\*  
## regulatoryQuality 2.163e-01 5.450e-02 3.969 8.04e-05 \*\*\*  
## GNIperCapita 5.901e-06 2.292e-06 2.574 0.01027 \*   
## HDI 4.968e-01 2.403e-01 2.067 0.03912 \*   
## polityScore 2.314e-02 1.027e-01 0.225 0.82183   
## democ 4.230e-03 1.048e-01 0.040 0.96782   
## autoc 1.895e-01 1.049e-01 1.807 0.07125 .   
## xrreg -1.273e-01 6.480e-02 -1.964 0.04995 \*   
## xrcomp -6.192e-02 6.445e-02 -0.961 0.33705   
## inverse\_pr 1.068e-01 3.828e-02 2.788 0.00545 \*\*   
## inverse\_cl 2.465e-01 4.199e-02 5.871 6.91e-09 \*\*\*  
## inverse\_mean NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for gaussian family taken to be 0.3752087)  
##   
## Null deviance: 611.39 on 659 degrees of freedom  
## Residual deviance: 243.51 on 649 degrees of freedom  
## AIC: 1238.9  
##   
## Number of Fisher Scoring iterations: 2

### *Predicting with Classification Tree*

In this classification tree, we are using “stability category” as the target variable. Using the variables identified through the above step function, we run machine learning predictions using classification tree with the caret package, and we obtain the following results:

* As shown in Figure 8, on the trainning data set, regulatoryQuality is the only variable that can predict either a case is “moderately unstable” (65%), “moderately stable” (22%), or “highly stable” (12%).
* As shown in Figure 9, on the test data set, two variables, civilLiberties and regulatoryQuality, are needed to predict either a case is “moderately unstable” (60%), “moderately stable” (29%), or “highly stable” (11%).

In sum, in both models, only regulatoryQuality is the most consistent predictor of “stability category”.

## n= 1986   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 1986 1296 Moderately Stable (0.11027190 0.19939577 0.34743202 0.34290030)   
## 2) regulatoryQuality>=0.410178 634 279 Moderately Stable (0.32649842 0.02523659 0.55993691 0.08832808) \*  
## 3) regulatoryQuality< 0.410178 1352 727 Moderately Unstable (0.00887574 0.28106509 0.24778107 0.46227811) \*

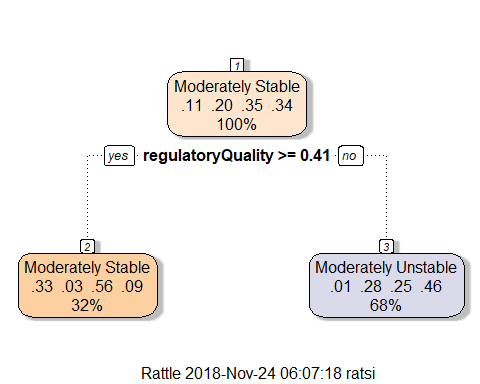


Fig. 8 - Classification Tree on the Trainning Data Set

## n= 661   
##   
## node), split, n, loss, yval, (yprob)  
## \* denotes terminal node  
##   
## 1) root 661 432 Moderately Stable (0.110438729 0.199697428 0.346444781 0.343419062)   
## 2) regulatoryQuality>=0.4430436 209 95 Moderately Stable (0.344497608 0.043062201 0.545454545 0.066985646)   
## 4) regulatoryQuality>=1.447581 62 17 Highly Stable (0.725806452 0.000000000 0.258064516 0.016129032) \*  
## 5) regulatoryQuality< 1.447581 147 49 Moderately Stable (0.183673469 0.061224490 0.666666667 0.088435374) \*  
## 3) regulatoryQuality< 0.4430436 452 239 Moderately Unstable (0.002212389 0.272123894 0.254424779 0.471238938) \*

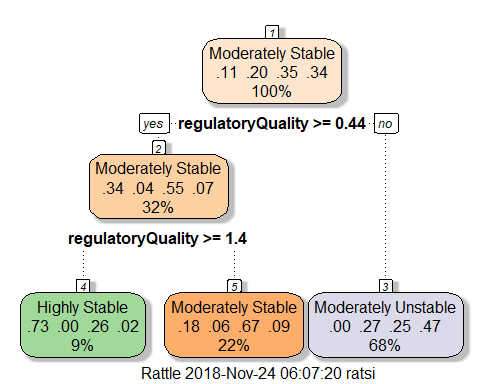


Fig. 9 - Classification Tree on the Testing Data Set

## References

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2011. “The Worldwide Governance Indicators: Methodology and Analytical Issues.” 1876-4053. Washington, DC: The World Bank.

World Bank. 2016. “Worldwide Governance Indicators.” *The World Bank Databank*. http://databank.worldbank.org/data/reports.aspx?source=worldwide-governance-indicators.