Third Pair Assignment

19 Apr 2016

Research Question

Our Collaborative Research Paper aims at shedding some light on what the effects of ethnic fractionalization on Wolgang Merkel's conceptualization of political rights are. In this sense, we have two main hypotheses:

H1: The more ethnically fractionalized a country is, the higher its negative impact on political rights if income inequality is high.

H2: When income inequality is low, the impact of high ethnic fractionalization on political rights is expected to be positive.

For more information regarding the theoretical underpinnings of our research proposal, you can have a look at our second pair assignment, which can be found here: Assignment 2

Model

Our aim is to collect data on level of political rights, ethnic fractionalization, and income inequality for all countries in the world during the period 2002-2014. We make use of Freedom House's freedom of the press indicator, drawn from the Quality of Governance dataset, as a proxy for political rights, our dependent variable (DV). Nevertheless, we also intend to make use of an organizational and associational freedom index as an alternative DV for a robustness check.

Our two main independent variables (IVs) of interest are ethnic fractionalization and income inequality (as an interaction variable). Other control variables that we include in our model are GDP per capita, judiciary independence, governmental accountability and linguistic fractionalization. Again, for more information concerning the theoretical reasons behind this model specification, please sese Assignment 2

Data sources

The variables that we are collecting for this research project come from three different data sources: the Quality of Government Standard Data, the World Development Indicators and the Worldwide Governance Indicators.

- Quality of Government (QoG) Standard Time-Series Data - These are contained in the QoG Standard Time-Series Dataset, which includes data from 1946 to 2015, the unit of analysis being country-year. It compiles approximately 2500 variables, making it the largest dataset developed by the Quality of Government Institute at the University of Gothenburg, Sweden.
- World Development Indicators (WDI) - These are a set of indicators developed by the World Bank, compiling important, high-quality, and internationally comparable statistics about global development and the fight against poverty. They are organized around six main themes, which are world view, people, environment, economy, states and markets, and global links.
- Worldwide Governance Indicators (WGI) - These are also developed by the World Bank. They consist of six composite indicators that measure broad dimensions of governance, covering approximately 200 countries since 1996: Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. They are obtained from 31 different perception-based data sources, aiming at enabling meaningful cross-country comparisons over time.

Data gathering

Quality of Government data

To collect the data from the Quality of Government Standard Time-Series Dataset (Quality of Government Institute 2015), we downloaded and imported the .csv file containing the data into R and dropped all variables but those we wish to include in our model for the period 2002-2014. These are freedom of the press and associational and organization freedom, both developed by Freedom House; indicators for ethnic and linguistic fractionalization (Alesina et al. 2003); and the World Economic Forum (WEF)'s judicial independence variable.

World Development Indicators

Using the WDI Package for R, we were able to use the World Bank indicator codes to automatically download the variables that we are interested in from the World Bank's APIs. These are: the Gini coefficient, which meassures income inequality (World Bank code: SI.POV.GINI), and GDP per capita in current US dollars (World Bank code: NY.GDP.PCAP.CD) (The World Bank 2014)

Worldwide Governance Indicators

Unlike in the case of the WDI, there is no API. Therefore, we had to download and import the data into R from a zipped .csv file. Since we were only interested in the Voice and Accountability variable for the years 2002-2014, we dropped the other five variables (Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption) as well as the values that the variable Voice and Accountability takes for previous years (1996, 1998, 2000) (Kaufmann, Kraay, and Mastruzzi 2010).

Cleaning and merging

After we downloaded the data and selected the variables that we want to include in our model, we cleaned and merged the QoG, WDI and WGI datasets. Our final dataset, FinalDataset.csv, has 2080 observations for 11 variables.

Descriptive and Inferential Statistics

Descriptive Statistics

To follow are some basic descriptive statistics examining our main variables of interest: Freedom of the Press (pressfreedom), Ethnic Fractionalization (ethnicfrac), and income inequality as measured by the Gini Coefficient (Ginicoef).

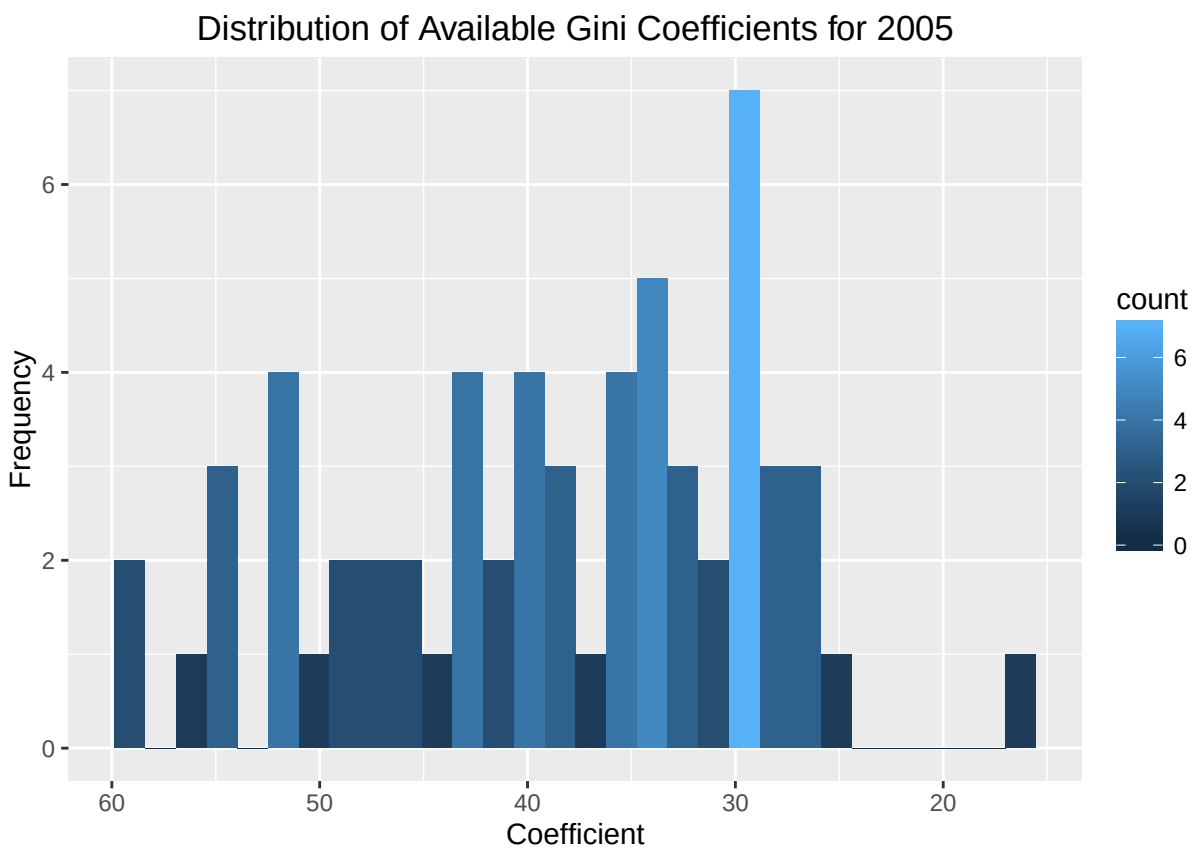
Measures of central tendency

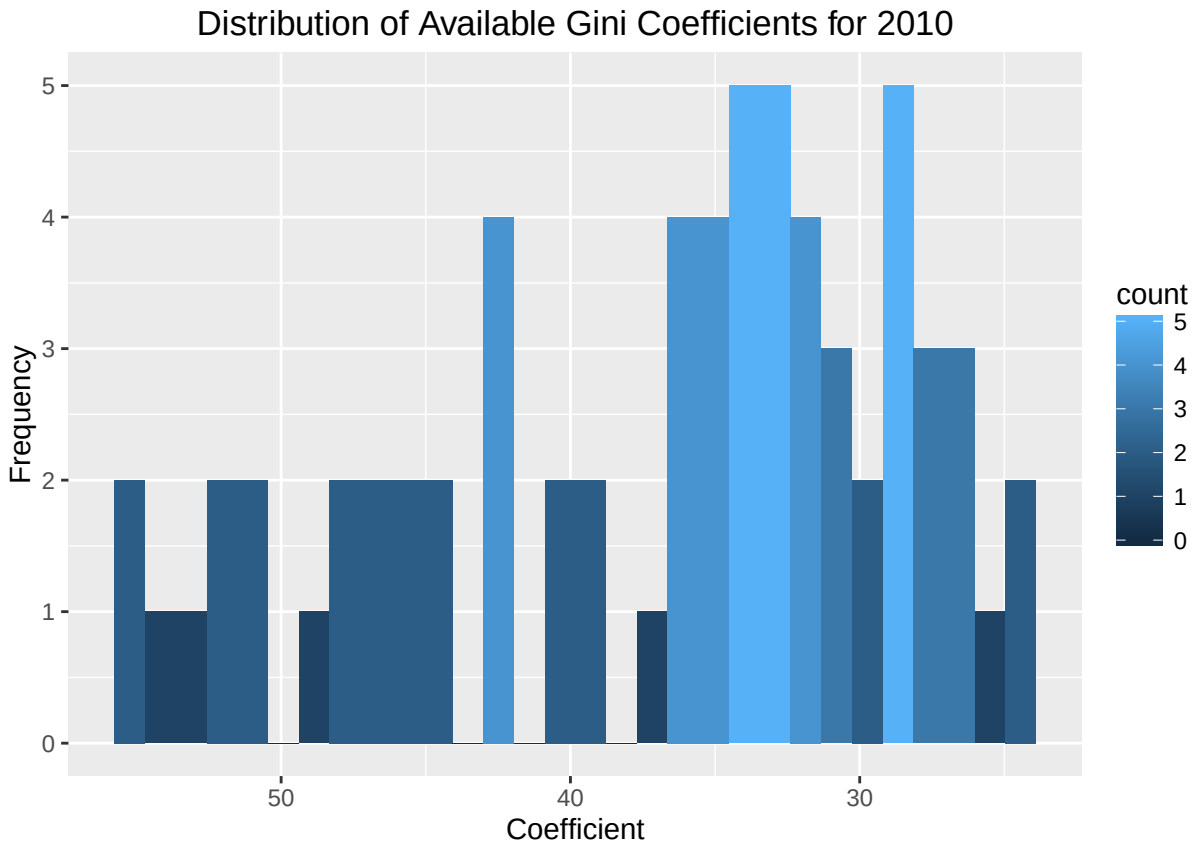
Table 1: Measures of Central Tendency

pressfreedom	ethnicfrac	Ginicoef
Min. : 8.00	Min. :0.0000	Min. :16.64
1st Qu.:23.00	1st Qu.:0.1866	1st Qu.:30.85
Median :44.00	Median :0.4173	Median :35.39
Mean :45.25	Mean :0.4325	Mean :38.10
3rd Qu.:64.00	3rd Qu.:0.6554	3rd Qu.:45.30
Max. :97.00	Max. :0.9302	Max. :64.79
NA	NA	NA's :1641

Histograms

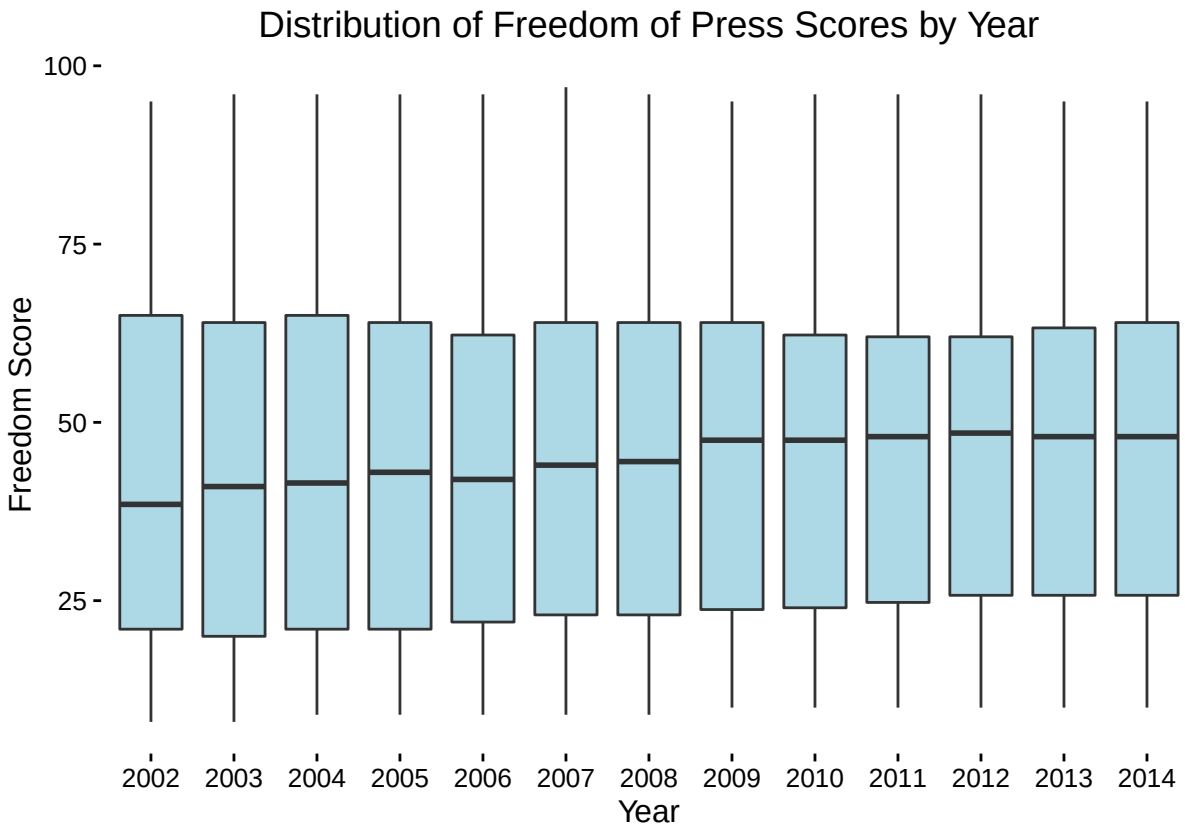
In the following figures, the data has been subset into years to examine changes over time. We can see the distribution of countries by relative income inequality for the years 2005 and 2010. These statistics are difficult to interpret, since while we can assume that income inequality does not change so dramatically in most countries over the course of 5 years, the graphs appear drastically different due to the inconsistency of data available from year to year. In other words, these two charts show different sets of countries; therefore, no conclusion can be drawn as to change in income inequality on a global scale over time.




Distribution of Available Gini Coefficients for 2010
Frequency
Coefficient
50
40
30
0
1
2
3
4
5
count
5
4
3
2
1
0

Boxplots

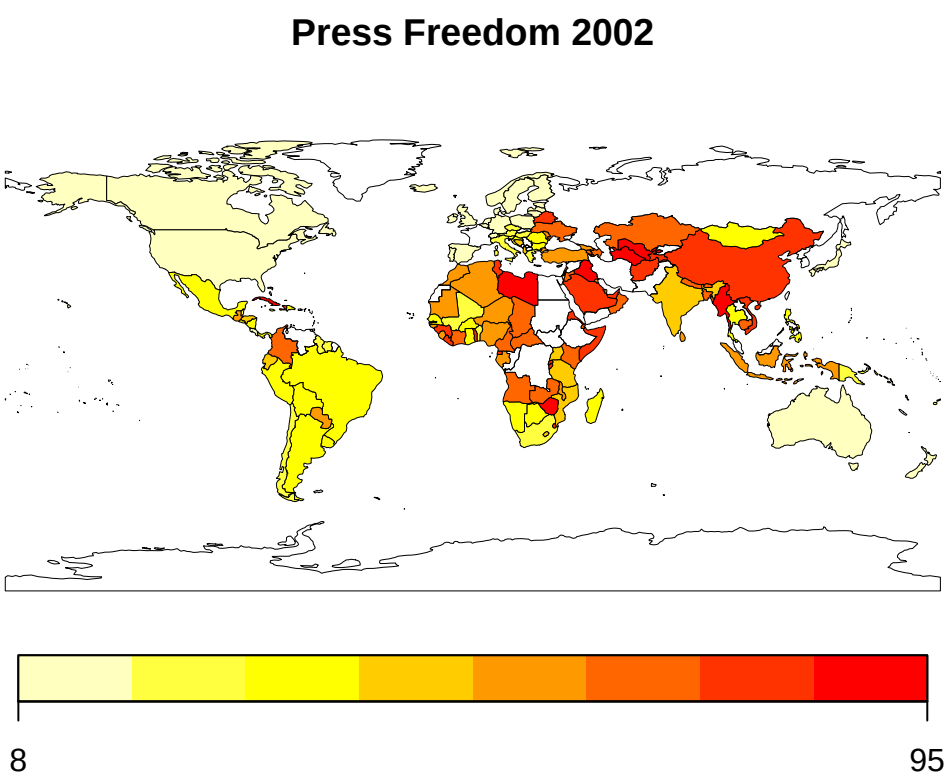
In the figure below, we can see the general distribution of Freedom of Press Scores among the world's countries by year. While the mean score does seem to be slightly increasing over time, there also seems to be a subtle movement from the edges inwards; that is to say, both countries that are on the upper end and the lower end of the spectrum are moving slightly towards the center. It is difficult to detect any significant patterns beyond this, however.



Maps

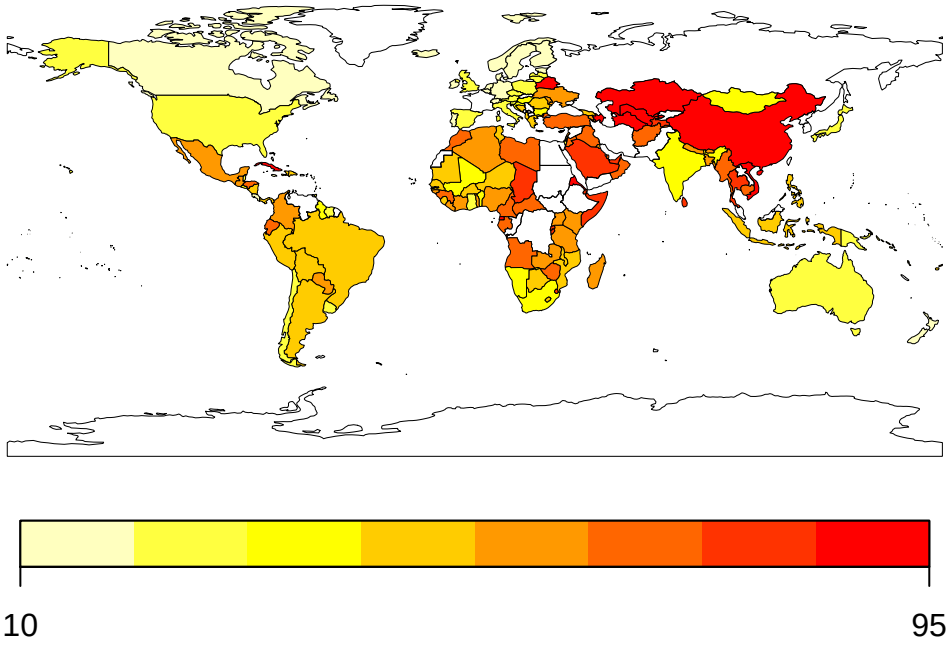
When we inspect press freedom on a map, on the other hand, it is interesting to see how quickly it can fluctuate over time with shifts in political regimes, governments, and societal conditions. The difference here from country to country is much more drastic than one might guess based on the box plot above.

```
## 160 codes from your data successfully matched countries in the map
## 0 codes from your data failed to match with a country code in the map
## 82 codes from the map weren't represented in your data
```

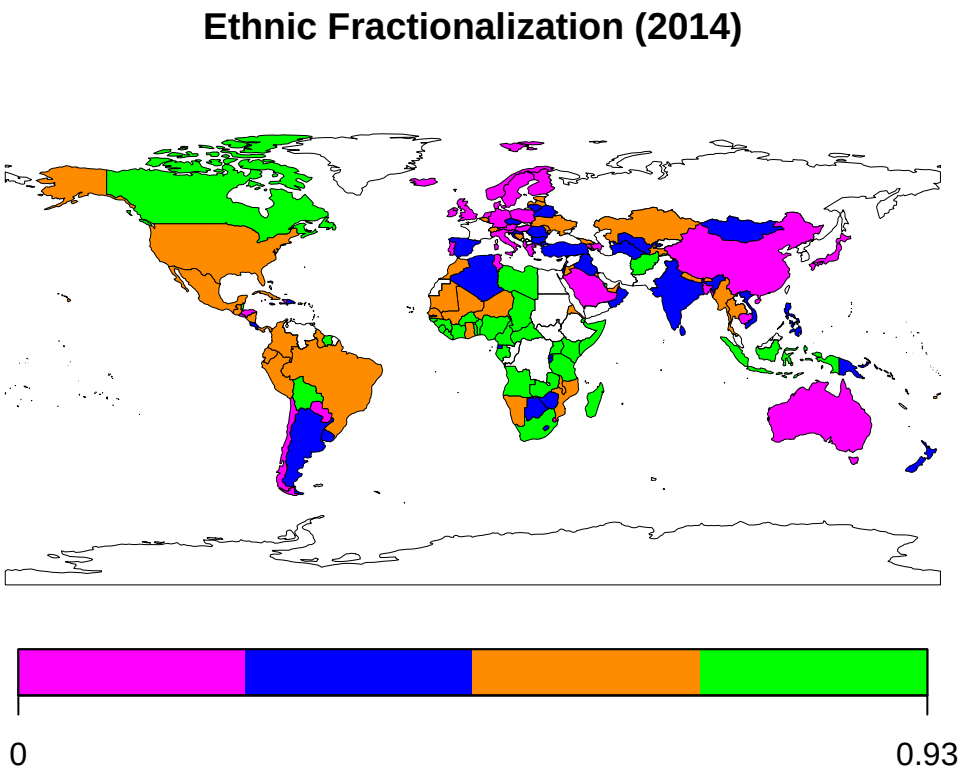



```
## 160 codes from your data successfully matched countries in the map
## 0 codes from your data failed to match with a country code in the map
## 82 codes from the map weren't represented in your data
```

Press Freedom 2014



Since ethnic fractionalization tends to remain relatively steady over time, we have chosen to look at the most recently available statistics. We have categorized countries into bins of relative ethnic fractionalization, where green is a high probability that any two individuals randomly selected from the population be of different ethnicities, magenta is a low probability (implying high ethnic homogeneity), and blue and orange fall somewhere in the middle. It is interesting to see how the upper half of the scale - those countries in green and orange - and the lower half - blue and magenta - cluster in regions of the world. While we believe this may have implications for similar patterns in strength of political rights, at first glance we see no comparable pattern between this map and the previous ones. The fact that this remains relatively more constant over time than press freedom also suggest that there are likely other factors we need to take into account.



Inferential Statistics

Although the inferential part of our analysis is just being started, we have already run an OLS and a fixed-effects models (the table with the coefficients for both models is in the next page). On the one hand, based on our preliminary OLS regression results, a 1-percent-point-increase in level of ethnic fractionalization is predicted to decrease freedom of the press by 6.99 points (let´s remember that each country is given a total press freedom score ranging from 0, best, to 100, worst). As for income inequality, a 1-point increase in the Gini coefficient is predicted to lead to an increase of freedom of the press by 0.12. Thus, its effect does not appear to be very significant from a substantive standpoint. Nevertheless, the effect of both main predictors is highly statistically significant. Regarding our control variables, in spite of the fact that only the level of linguistic fractionalization of a country is not statistically significant, the voice and accountability variable clearly appears to be the most substantively significant, since a 1-point increase in the World Bank index is predicted to improve a country´s press freedom score by 23.35 points (thus, the press freedom score given to a country is expected to decrease by 23.35 points).

On the other hand, based on our fixed-effects model, only income inequality, as measured by the Gini coefficient, and the World Bank´s voice and accountability variable appear to hold some explanatory power (statistically significant) when it comes to making sense of the differences in press freedom scores across countries and years. As in the OLS model, the effect of voice and accountability is the most prominent from a substantive standpoint (a 1-point increase in voice and accountability is expected to improve a country´s press freedom score by 12.52). Concerning our main IV, the level ethnic fractionalization, as well as the level of linguistic fractionalization variable, as they vary very little across time (they are nearly time-invariant), both of them drop out when running a fixed-effects model.

Table 2: Regression Results: Pooled OLS (1) and Fixed Effects (3)

	Dependent variable:	
	Freedom of the press	
	(1)	(2)
Ethnic Fractionalization	-6.99^{***}	
	(2.00)	
Gini coefficient	0.12^{***}	-0.29^{***}
	(0.04)	(0.08)
GDP per capita	0.0001^{***}	0.0000
	(0.0000)	(0.0000)
Voice and Accountability	-23.35^{***}	-12.52^{***}
	(0.60)	(2.26)
Judicial Independence	-1.34^{***}	-0.01
	(0.34)	(0.49)
Linguistic Fractionalization	1.53	
	(1.61)	
Intercept	49.39^{***}	
	(1.97)	
Observations	341	341
R^2	0.93	0.18
Adjusted R^2	0.91	0.13
F Statistic	712.65^{***} (df = 6; 334)	13.43^{***} (df = 4; 246)
Note:		*p<0.1; **p<0.05; ***p<0.01

Bibliography

Alesina, Alberto, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. 2003. "Fractionalization." http://www.nber.org/papers/w9411.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2010. "The Worldwide Governance Indicators: Methodological and Analytical Issues." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682130.

Quality of Government Institute. 2015. "QoG Standard Data." http://qog.pol.gu.se/data/datadownloads/qogstandarddata.

The World Bank. 2014. "World Development Indicators." http://data.worldbank.org/sites/default/files/wdi-2014-book.pdf.