Predicting Sovereign Credit Ratings with Economic, Social, and Political Factor

Final Project: Data Science II (STAT 301-2)

Irene Kwon

Introduction

The question of why nations keep their promises – be it contracts or laws – has long intrigued scholars in International Relations. Without higher authority to enforce the obligations, why do countries follow contractual obligations? Sovereign debt is no exception; it is a contract between a government and a foreign lender (often in private sector), stating an obligation for sovereigns to pay back its outstanding liabilities within a fixed period of time. Then how can lenders ensure that borrower governments will pay back their obligations?

The role of credit rating agencies is hence important in sovereign debt markets. Their ratings provide a standard to which market participants can anchor their expectations on a country's willingness and capacity to pay back its loan. In other words, these rating agencies evaluate and codify not only how capable a country is to repay its debts, but also their willingness to do so.

This leads to my research questions: how can we *predict* major CRAs' sovereign ratings? Since CRAs have to evaluate the *willingness* of a country to repay their debts, do political factors matter more in predicting the sovereign ratings? Or instead, will economic variables suffice to predict the creditworthiness of a country?

My final project is an extension from the last EDA project on credit rating agencies' sovereign ratings. While my previous EDA project provided an extensive vignette of the sovereign debt ratings for 115 countries of 30 years, I now focus mainly on high-income countries.

First, since the high-income countries already pass an income threshold, their capacity to pay back might be less of an issue to credit rating agencies. Rather, we can expect that other factors – societal and political – might better predict a country's sovereign ratings than its income level. Another explanation could be that after the 2008 Great Recession and 2010 Eurozone Crisis, fiscal and economic soundness, rather than the GDP per capita (i.e., the level of wealth per se), might be weighted more heavily.

Sovereign ratings determine the terms of loans and the access to the international financial market. Rich countries are no exceptions – they are subject to the ratings, and their loan conditions are also dependent upon these ratings. Hence, the prediction of the sovereign ratings is a practically and intellectually important research agenda.

Using the models we covered in this class, I examine which variables predict sovereign ratings among high-income countries the best. The report proceeds as follows. First, I start with a brief introduction of the data set. This is to provide an overview of what the data contain and clarify the response variable as well as potential predictors. Second, I present an exploratory data analysis (EDA) to familiarize the readers with the data set. Not only does EDA help us understand the structure of the data, but the preliminary findings from this step can be used as a foundation for model building. Third, I set up a series of plausible models. After model construction, I select the best predictive model(s) using k-fold cross-validation (where k = 10). Lastly, as my final step, I will test the performance of the models that "won" in the k-fold cross-validation process on the test set. I conclude by providing a brief summary of the findings and some suggestions for further works.

Data Overview

1. Data Description

I gratefully received the sovereign rating datasets from Professor Pedro Gomes at Birkbeck University of London. His dataset compiled sovereign debt ratings from 1989 to 2006, which I expanded to include the ratings up till 2018.

The original dataset from the last quarter contained 3,450 observations for 24 substantive economic and political variables. The included observations record 115 countries for 30 years (from 1989 to 2018). Since I am interested in the prediction of sovereign ratings among "rich" countries, I only include high-income countries per World Bank's income classification. This leaves 1,132 observations in my scope of analysis. The basic unit of analysis is country-year, with each row containing a given year's observation for a country for each given variable.

- **Main Response Variable**

Per my research question, the main response variable is the credit rating agencies' sovereign ratings. The dataset contains three agencies' ratings: Moody's, Fitch, and S&P. As I show in the EDA section later in this report, the three rating agencies "agree" a lot when it comes to the issuance of sovereign ratings. Hence, I combine the ratings by averaging them to create one main response variable.

The response variable is coded in 17-point scale, with 17 being the highest rating and 1 being the lowest. Since this 17-point scale is granular enough, I treat this as a continuous, numerical variable. Therefore, this project concerns a *regression* problem rather than a *classification* problem.

- **Predictor Variables**

As potentially important predictors for sovereign credit ratings issued by American rating agencies, I include a wide range of political, institutional, and economic predictors.

Table 1: Data summary

Name	Piped data
Number of rows	1132
Number of columns	35
Column type frequency:	
factor	3
numeric	32
Group variables	None

Variable type: factor

skim_variable	n_missing	complete_rate	ordered	n_unique	top_counts
country	0	1.00	FALSE	52	Aus: 30, Aus: 30, Bah: 30, Bel: 30
voc	591	0.48	FALSE	4	CME: 240, LME: 210, SME: 81, HME: 10
region	0	1.00	FALSE	8	Wes: 420, Mid: 197, Sou: 138, Eas: 98

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p2
no	0	1.00	1.741230e+03	1.055060e+03	2.600000e+01	9.237500e+0
code	0	1.00	5.847000e+01	3.516000e+01	1.000000e+00	3.100000e+0

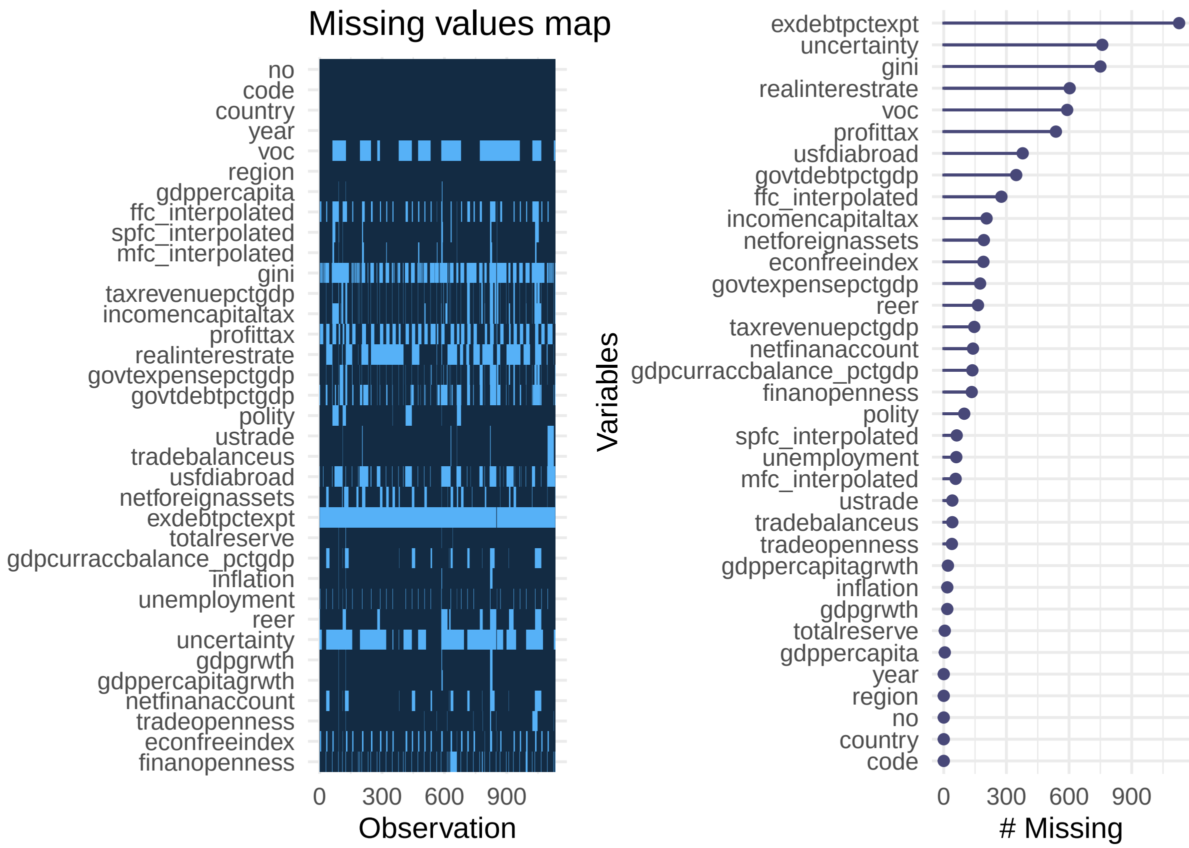
skim_variable	n_missing	complete_rate	mean	sd	p0	p2
year	0	1.00	2.005210e+03	8.610000e+00	1.989000e+03	1.998000e+0
gdppercapita	5	1.00	3.201068e+04	1.822394e+04	5.419590e+03	1.880784e+0
ffc_interpolated	276	0.76	1.415000e+01	3.310000e+00	1.000000e+00	1.200000e+0
spfc_interpolated	62	0.95	1.404000e+01	3.390000e+00	1.000000e+00	1.200000e+0
mfc_interpolated	57	0.95	1.407000e+01	3.500000e+00	1.000000e+00	1.200000e+0
gini	750	0.34	3.173000e+01	4.380000e+00	2.370000e+01	2.810000e+0
taxrevenuepctgdp	146	0.87	1.957000e+01	8.380000e+00	4.000000e-02	1.366000e+0
incomencapitaltax	205	0.82	2.916000e+01	1.471000e+01	3.400000e-01	1.872000e+0
profittax	537	0.53	1.330000e+01	9.080000e+00	-2.000000e-01	6.100000e+0
realinterestrate	603	0.47	4.230000e+00	6.580000e+00	-3.791000e+01	1.980000e+0
govtexpensepctgdp	174	0.85	3.356000e+01	1.353000e+01	3.900000e+00	2.473000e+0
govtdebtpctgdp	347	0.69	6.898000e+01	4.032000e+01	1.890000e+00	3.890000e+0
polity	98	0.91	6.910000e+00	6.400000e+00	-1.000000e+01	9.000000e+0
ustrade	41	0.96	3.884862e+04	8.456237e+04	1.155600e+02	2.627620e+0
tradebalanceus	41	0.96	-5.327260e+03	1.706286e+04	-1.003911e+05	-5.043900e+0
usfdiabroad	378	0.67	5.280210e+03	1.115765e+04	-4.136581e+04	1.713700e+0
netforeignassets	192	0.83	6.638045e+12	3.695681e+13	-3.050000e+13	-1.529020e+0
exdebtpctexpt	1127	0.00	1.266800e+02	4.223000e+01	9.014000e+01	9.283000e+0
totalreserve	5	1.00	7.536663e+10	1.578439e+11	1.199771e+08	6.314767e+0
gdpcurraccbalance_pctgdp	137	0.88	1.520000e+00	1.135000e+01	-2.405200e+02	-3.050000e+0
inflation	17	0.98	2.790000e+00	5.410000e+00	-2.763000e+01	9.400000e-0
unemployment	60	0.95	6.800000e+00	4.150000e+00	1.400000e-01	3.950000e+0
reer	164	0.86	1.019600e+02	1.268000e+01	6.757000e+01	9.562000e+0
uncertainty	759	0.33	1.194200e+02	5.695000e+01	3.760000e+01	8.252000e+0
gdpgrwth	17	0.98	2.800000e+00	3.540000e+00	-1.472000e+01	1.270000e+0
gdppercapitagrwth	20	0.98	1.510000e+00	3.190000e+00	-1.515000e+01	1.900000e-0
netfinanaccount	140	0.88	-3.215637e+09	9.000499e+10	-8.810000e+11	-6.514294e+0
tradeopenness	39	0.97	9.995000e+01	7.090000e+01	1.601000e+01	5.863000e+0
econfreeindex	190	0.83	7.043000e+01	7.020000e+00	3.630000e+01	6.560000e+0
finanopenness	134	0.88	1.730000e+00	1.120000e+00	-1.910000e+00	1.290000e+0

As can be seen from the summary above, most of the variables are continuous, numerical variables. There are only three categorical variables - `country`, `voc` and `region`.

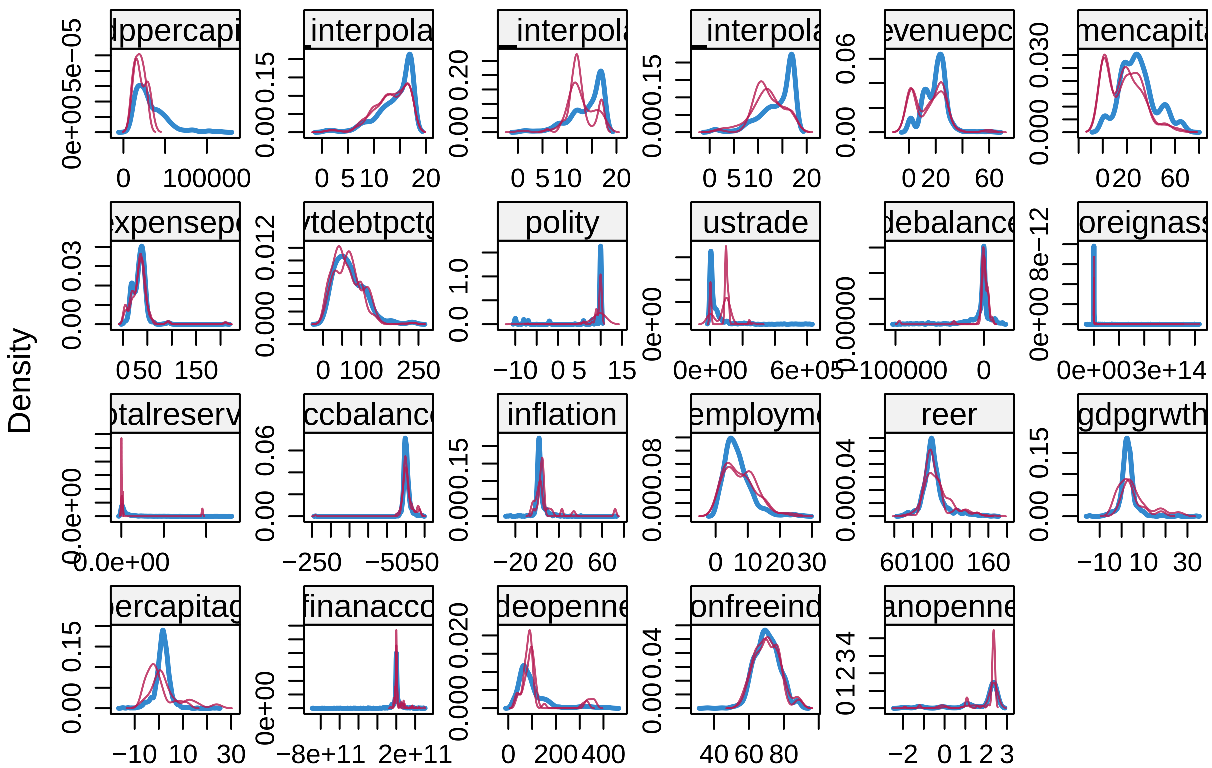
2. Dealing with Missingness

The *population*, or the target group of countries I am interested in exploring for this project is high-income countries. Since the target population only comprises high-income countries - meaning that the observations (i.e., countries) included in the dataset are at least relatively homogeneous in terms of GDP per capita - I anticipate multiple imputations to be less problematic to "fill in" the missing values. These high-income countries did not suffer from the *external shocks* such as wars, which are likely to have affected the missingness in the dataset.

First, the following plot reveals the missing patterns in the dataset.



As we've seen above from the skim results above, some variables have more missing values than others. For example, most of external debt as percent of export (`exdebtpctexpt`) is missing; while total reserve is almost complete. To enhance the accuracy/performance of multiple imputation, I only include variables with higher than 80% of the completion rates. This excludes `gini`, `profittax`, `realinterestrate`, `usfdiabroad`, `exdebtpctexpt`, and `uncertainty`. I use `cart` (classification and regression tree) method for multiple imputation; I generate two sets of multiply imputed datasets from 50 iterations of imputations for each set. One set will be used for data exploration (EDA), and the other set will be used for model building and model selection. In the Appendix, I check whether the two sets are similar.



The graph above shows the distribution of the newly imputed values in comparison with the existing values. The intuition is that the red and blue lines should not be that much different from each other. Although we see some deviations of the two lines, the overall distribution seems to be similar, indicating that I can use the imputed results for analysis.

3. Data Split

To maintain that we are using as "intact" data as possible for each part of our analysis (EDA, model building and model evaluation/validation), I will split up the data set.

First, as stated above, out of the two imputed sets, I will use one iteration of multiply imputed data for EDA and the other set for model building and validation. The assumption is that the two imputation results are relatively similar to each other - and I check this assumption in the Appendix.

Second, due to the panel structure of the data, the data set has a time-dependence problem. Therefore, per our previous lab on stock market data, I split the training and test (validation) set according to the time. To be specific, I set aside the observations of the years 2016-2018 as validation set. This has another advantage of answering one theoretically interesting question: in this way, we are examining how the models built on the previous time series can be used to predict the sovereign ratings of the future.

EDA

1. Panel Structure of the Data

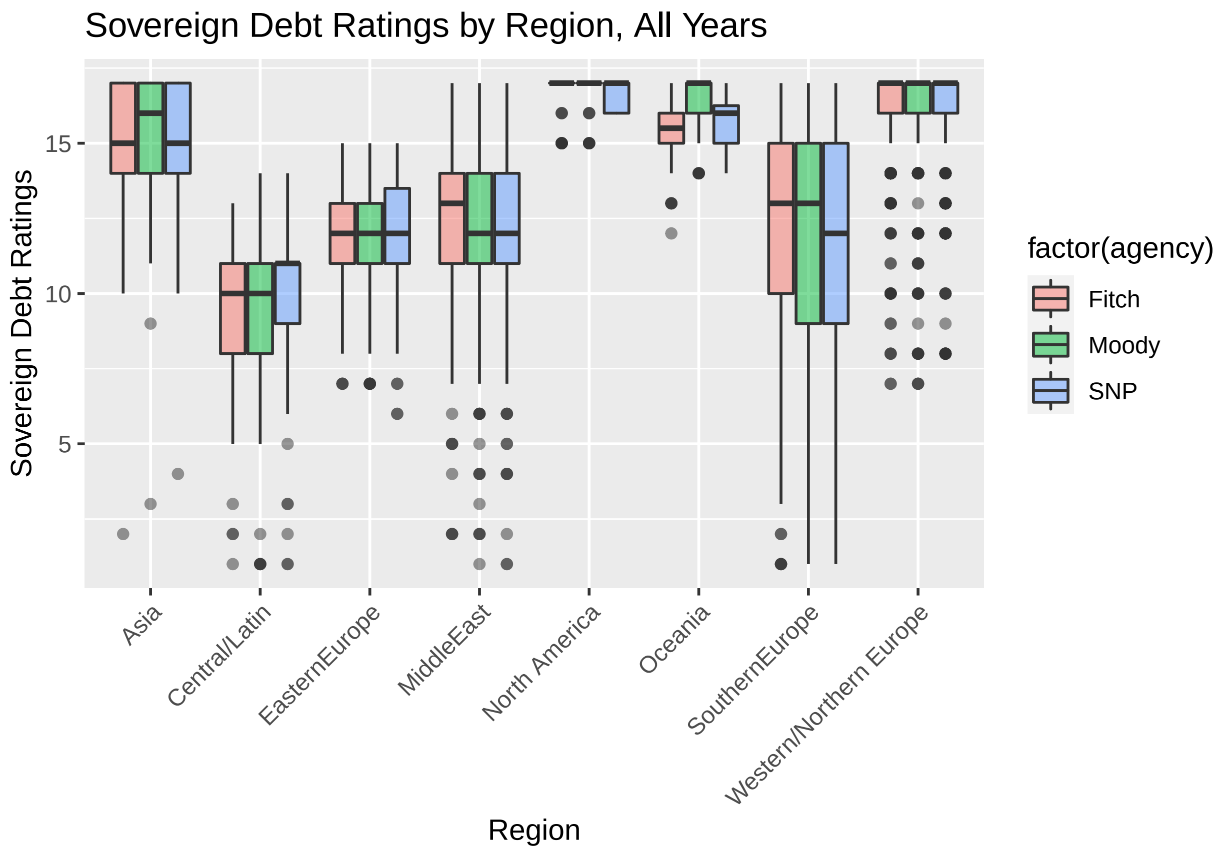
As mentioned above, the dataset has a panel structure, with the basic unit of analysis being country-year. Before proceeding to the actual exploration of the dataset - e.g., examining the correlation between the variables, looking at the distribution of the variables, etc. - I first check whether there are "anomalies" in the dataset. This does not necessarily mean outliers; I check whether we should remove some of the *observations* (rather than *variables*) that simply prevents me from running the statistical models. Spoiler alert: I did have to remove some countries!

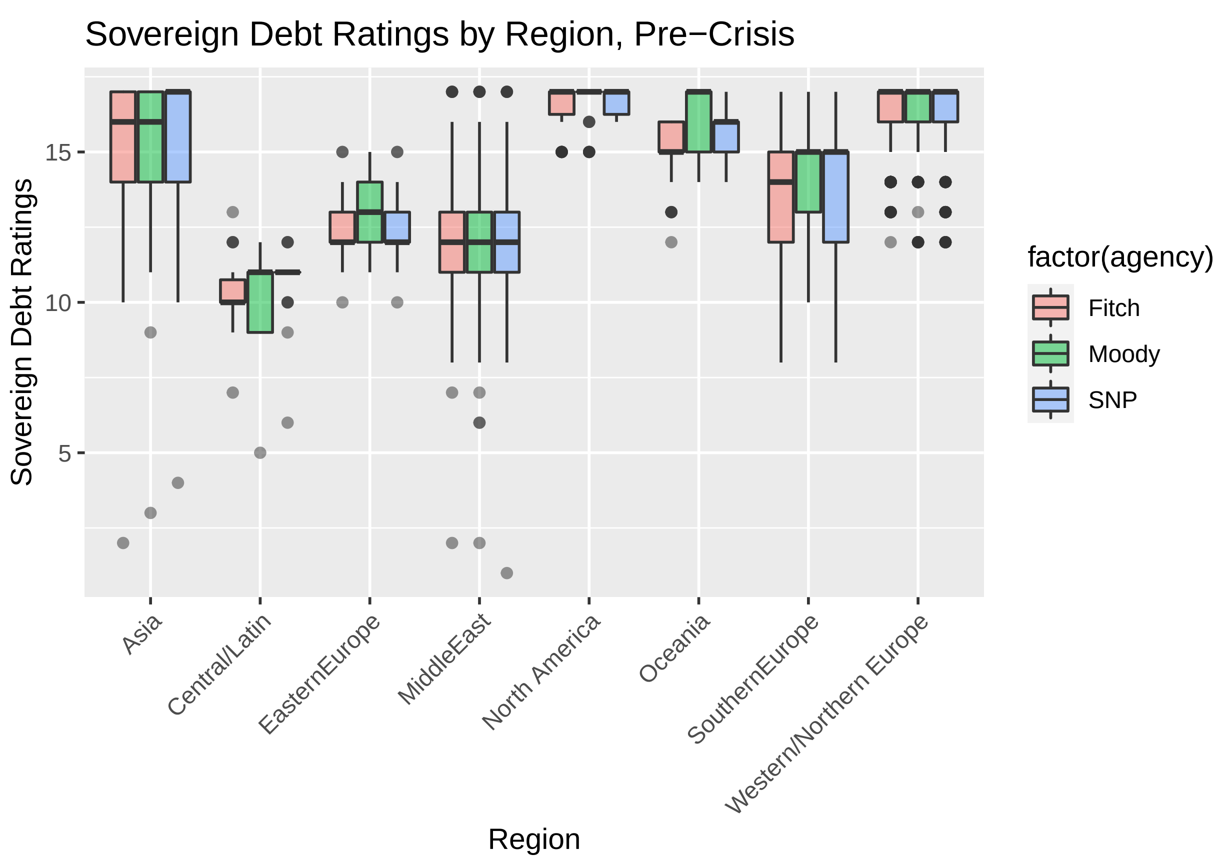
country	year	gdppercapita	mfc_interpolated	ffc_interpolated	spfc_interpolated
Argentina	2014	12334.80	1	1	1
Argentina	2017	14591.86	3	3	4
Australia	1989	17798.07	15	15	15
Australia	1990	18211.14	15	15	15
Australia	1991	18821.25	15	15	15
Australia	1992	18569.51	17	15	15
Australia	1993	17633.98	17	15	15
Australia	1994	18045.87	17	15	15
Australia	1995	20319.63	17	14	15
Australia	1996	21861.33	17	15	15

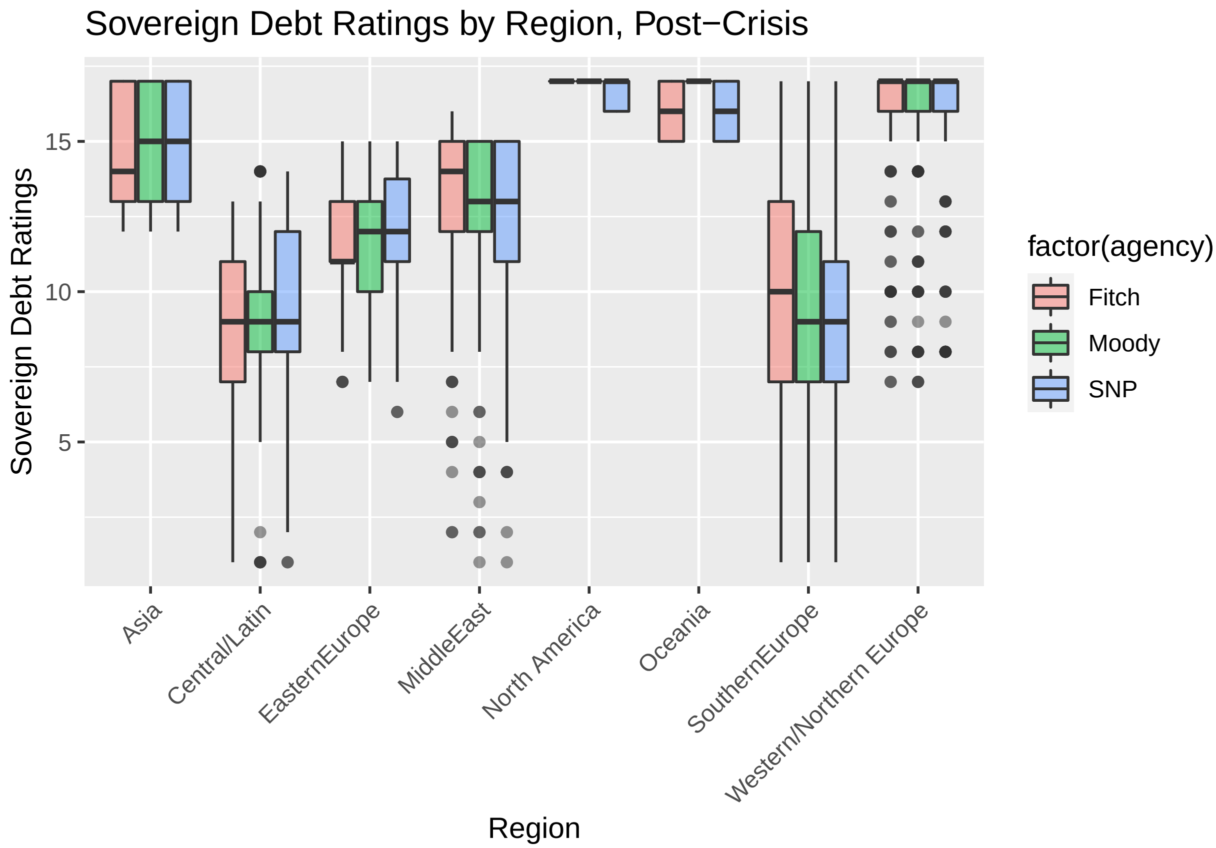
One noticeable thing about the content of this dataset is that it is an unbalanced panel. Since I limit the scope of analysis to only high-income countries - i.e., those which already met certain income levels - there are countries like Argentina which did not make it to the list in all 30 years covered in the dataset. Argentina met the threshold for high-income countries in the years 2014 and 2017, but not in any other years between 1989 and 2018.

Including such countries in the analysis can be problematic, since I plan to use country-fixed effects and lagged variables in the models; (1) they will be removed from the analysis anyways, or (2) because of their lack of variations (within the country), it might skew the results. Therefore, I remove the countries with less than five observations - i.e., those which did not meet the high-income threshold for more than five years. Such countries include Argentina, Panama, Russia and Venezuela.

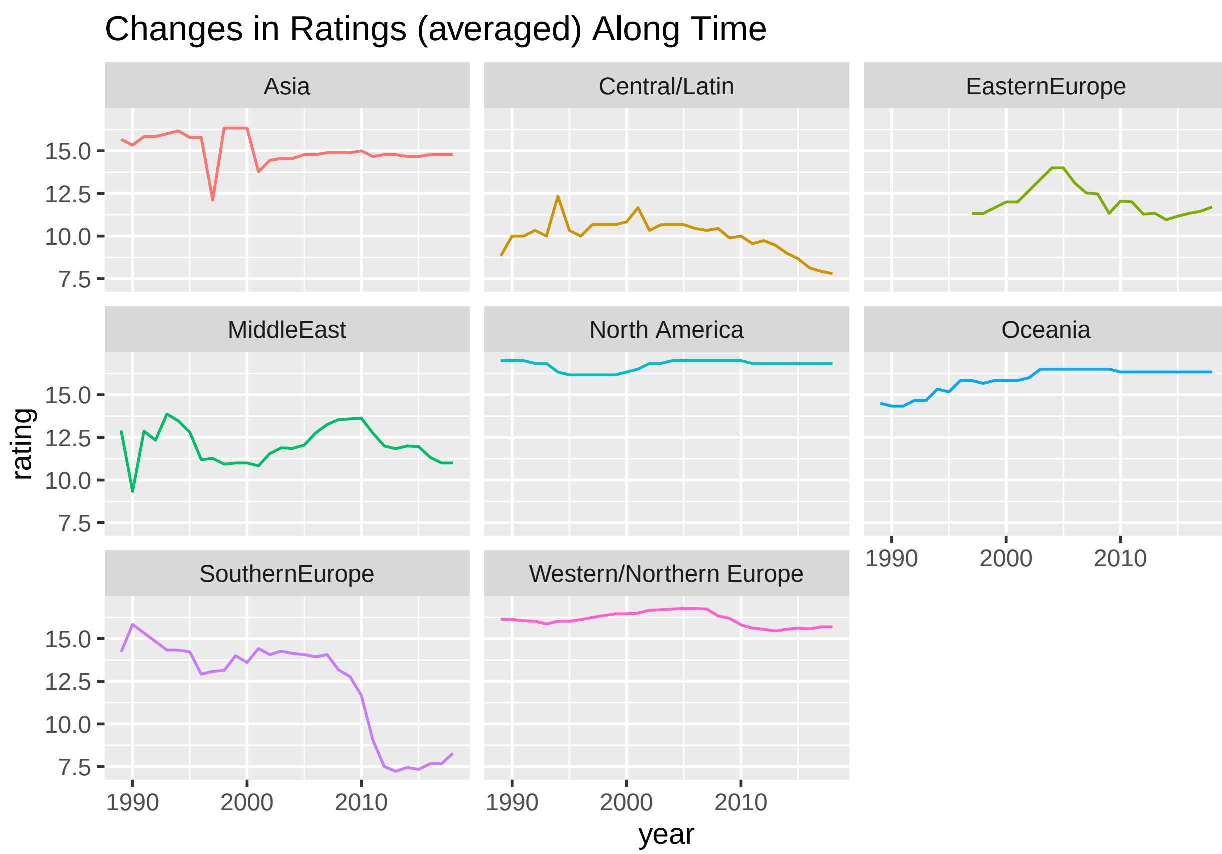
2. Graphic Representation of the CRA ratings (by region, by VoC)







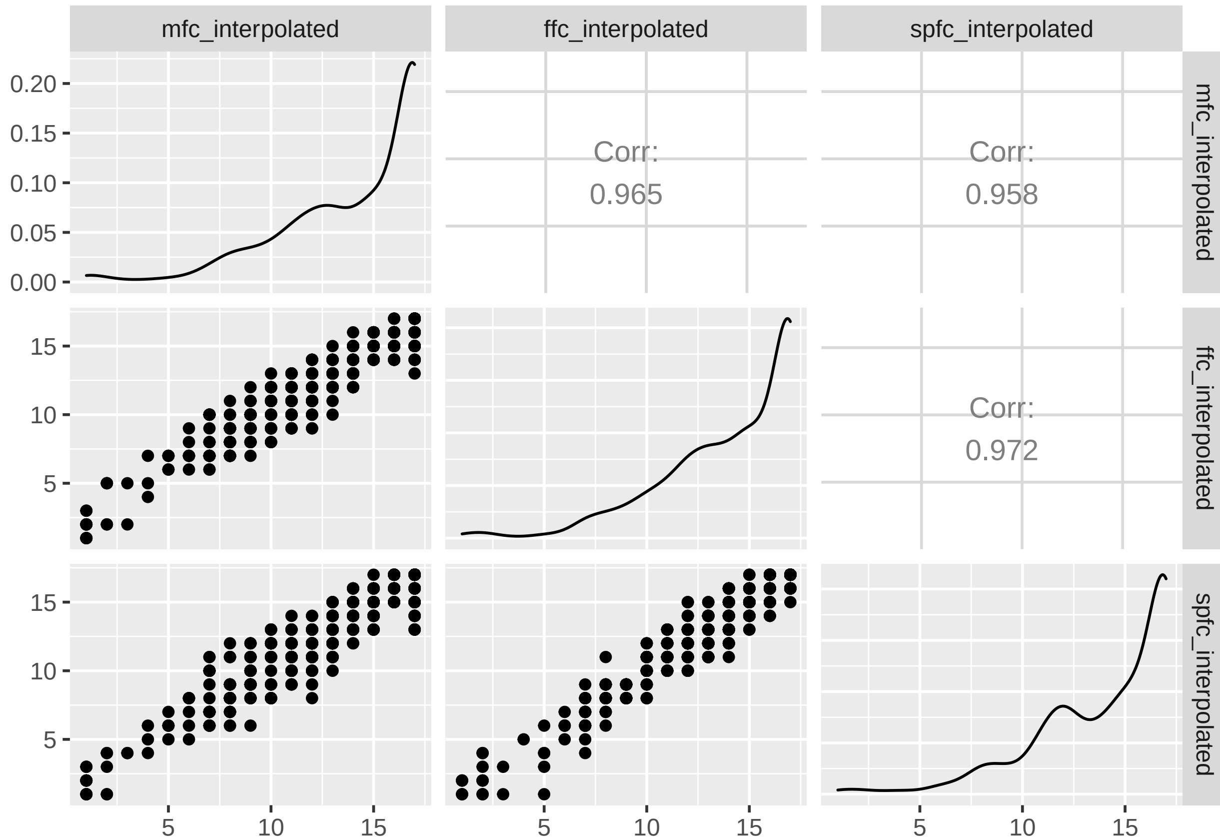
We can see that even within the high-income countries, there are regional variations in sovereign debt ratings. North America, Oceania, and Western/Northern Europe all did better than other regions during the 30-year period covered in the dataset, while Central/Latin American countries had the lowest ratings on average. Southern European countries had huge variations probably due to the plummeting during the Eurozone Crisis of late 2000s and early 2010s. We can see the changes in the general ratings in different regions below:



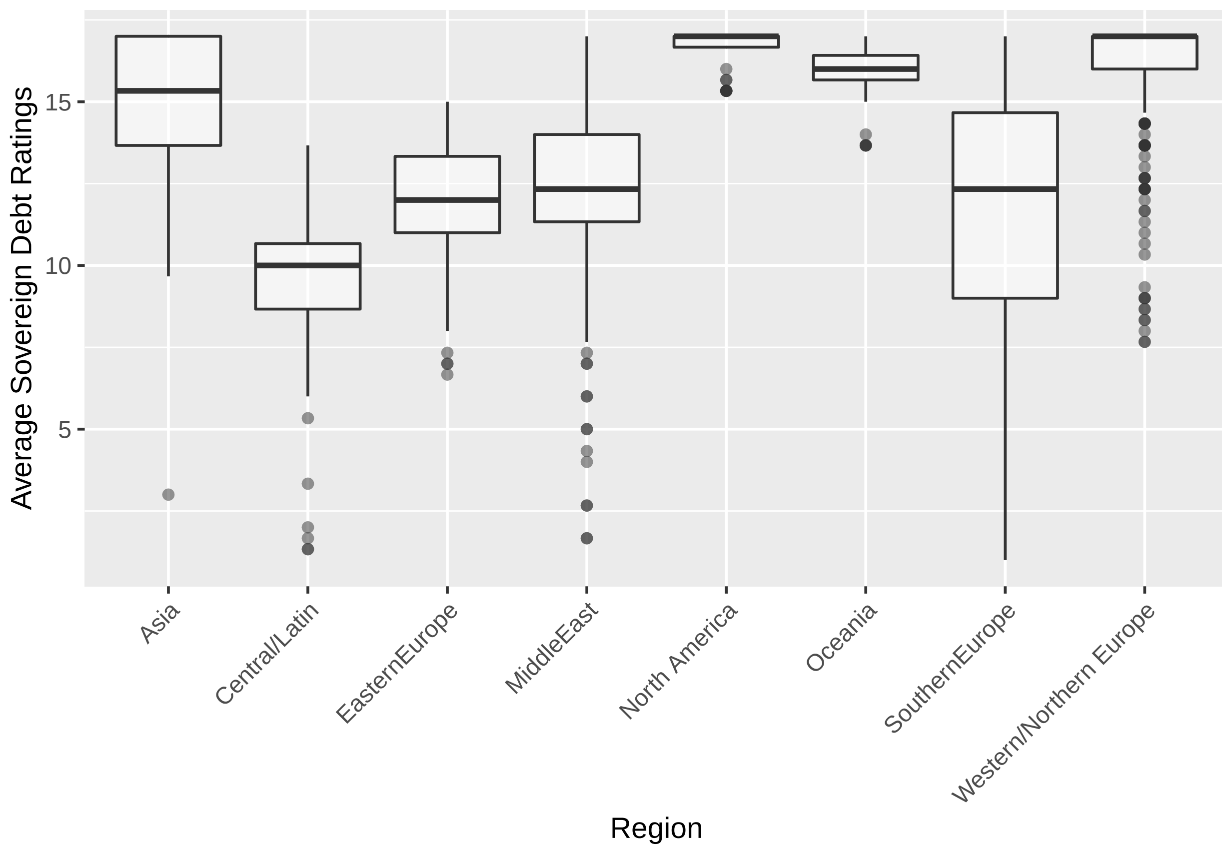
3. Correlation between Raters

Theoretically, we can run the same models on different agencies' ratings separately, three times, and compare the results. However, if the ratings from three agencies are relatively similar to each other, it might be more efficient if we could run just one model. Therefore, if each agency's ratings are generally similar to those of others, then I could combine the ratings and run predictive models on (only) the newly created response variable instead (I have tried running the identical models on different agencies' ratings separately for another project - with different research questions - for my previous conference paper, and the results were relatively similar).

Hence, I create a new variable, `rating`, by taking the average of the three agencies' ratings. Using the average variable has another advantage of 'averaging out' the marginal differences between the raters. For the analysis, I use `rating` as the main response variable.



As can be seen from the bivariate scatterplots and correlation coefficients, the ratings from different agencies tend to go hand in hand. Therefore, I combine the three ratings (by different agencies) into one average score. Following is the distribution of the combined/averaged ratings by region:



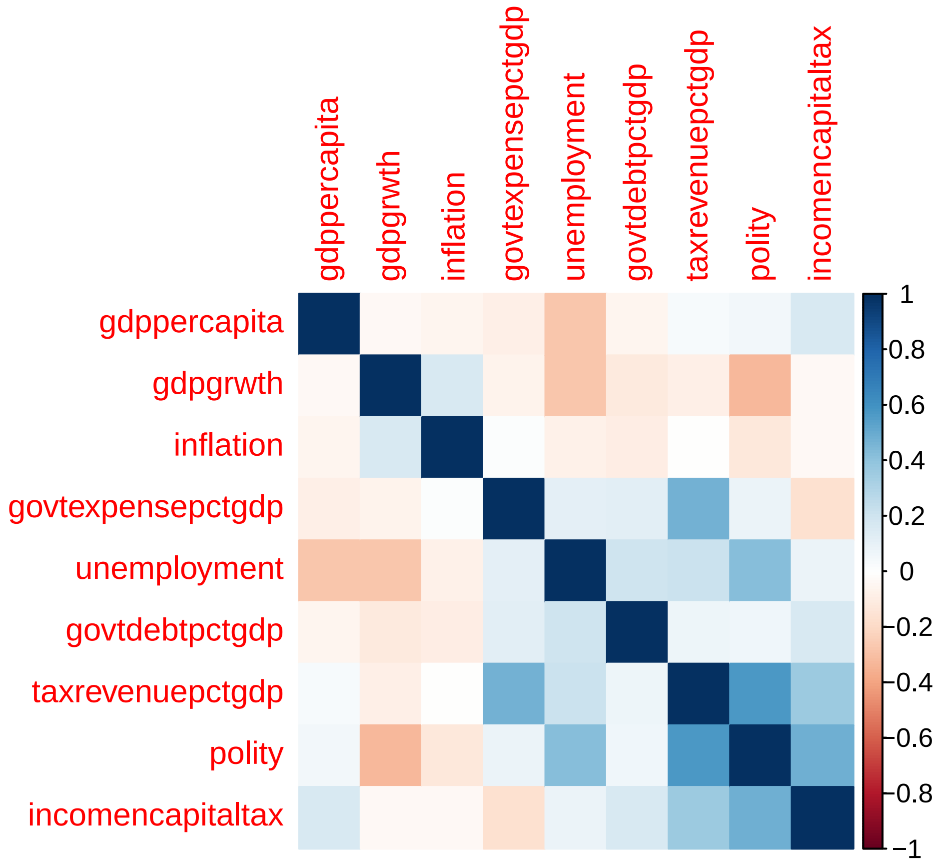
This boxplot looks exactly like what we'd expect to see based on the previous boxplots about each agency's ratings, suggesting that the new variable is a viable measure to use. In sum, I will use the average of the three ratings as the main response variable for my analysis.

4. Correlation between Predictors

The goal of this section is to check whether there are correlations between the predictors. It allows us to spot potential multicollinearity in the model, though this is a prediction-oriented project and multicollinearity would not be a big problem. Nevertheless, grasping what is going on between the predictors can be helpful in that it allows which variables - if any - to remove from the model. If a certain subset of variables are strongly correlated with each other, then highly likely are they indexing the same thing, though with different metrics and measures. In such a case, by removing a "duplicate" variable, we can keep the model simple and avoid potential overfitting.

(1) Correlation between Domestic Fiscal and Political Factors

- **correlation heatmap**



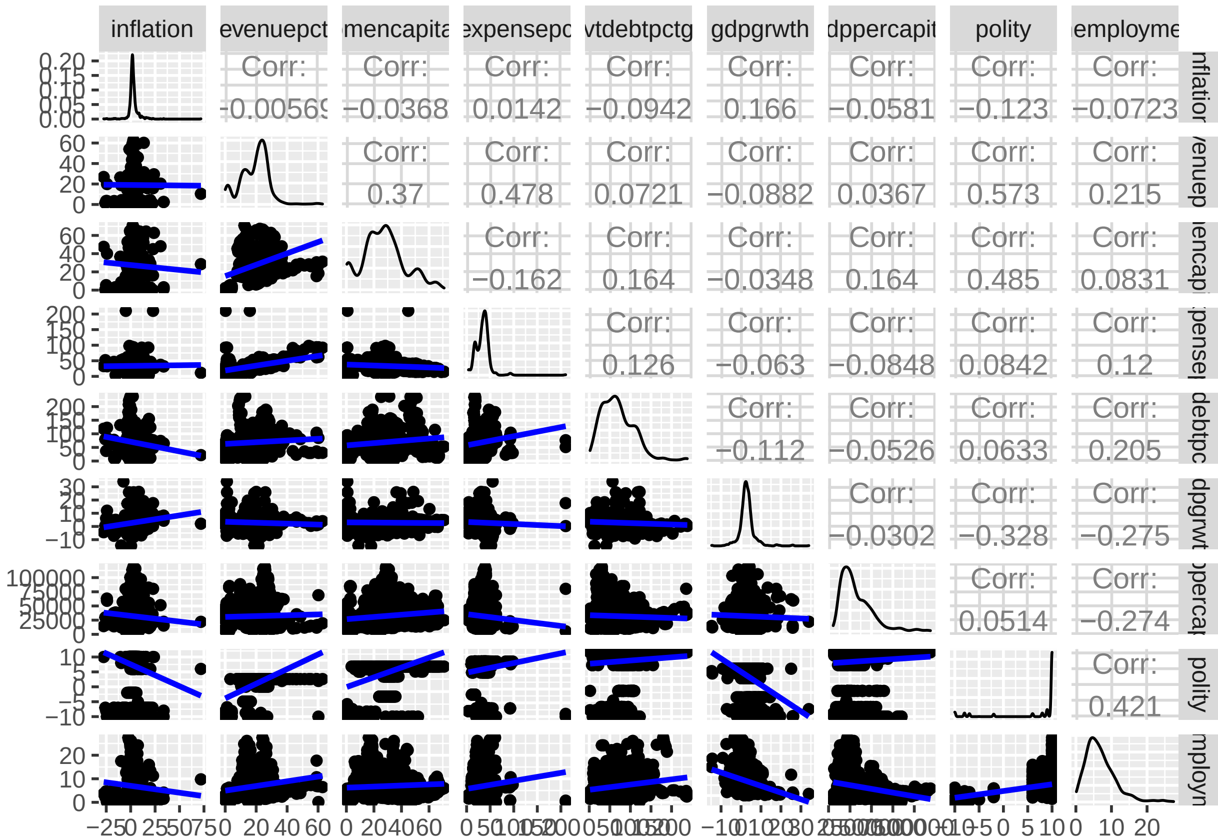
From the correlation heatmap, we can discover some interesting correlations between predictor variables. First, it seems that there is some positive association between tax revenue (measured as percent of GDP) and government expenses (also measured as a percent of GDP).

Second, we can also discover the positive correlation between the Polity Score and tax revenues. This makes sense, since one core element of the internal sovereignty includes the government's (or state authority's) capacity to tax its people. As the level/degree of the rule of law, or level of democracy, got higher, tax revenue as percent of GDP also got higher - probably meaning that there is some association between "solid" governance and taxation, at least among high-income countries.

Third, we can also discover the negative association between unemployment and GDP per capita, as well as that between unemployment and GDP growth. As unemployment itself signals slow economy, this negative correlation also makes sense.

Fourth, there is a positive correlation between unemployment and polity. This is something I could not expect! On the other hand, unemployment and inflation did not have high correlation - which contradicts what the Phillips Curve would suggest.

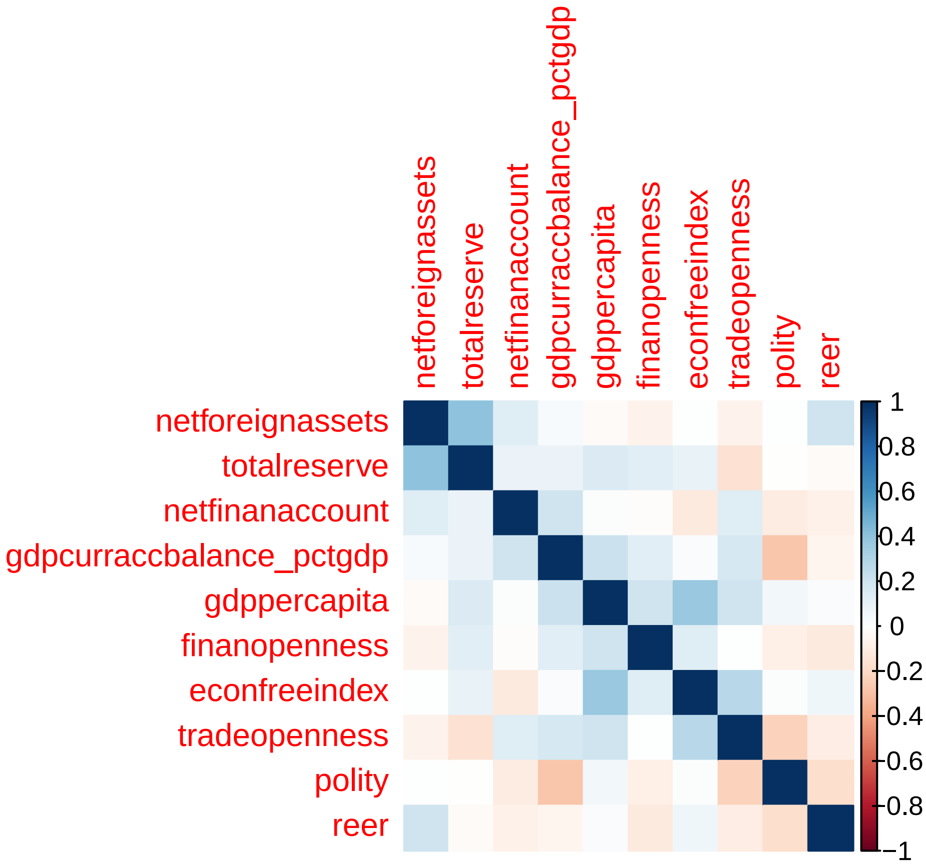
- **bivariate scatterplots**



Again, the pairwise scatterplots corroborate our findings from the heatmap. Though some correlation coefficients are as high as 0.48, I do not spot serious multicollinearity issue from the bivariate correlations at least. It seems that previous theories and literature did a good job pointing out relevant variables!

(2) Economic Structure and International Economic Factors

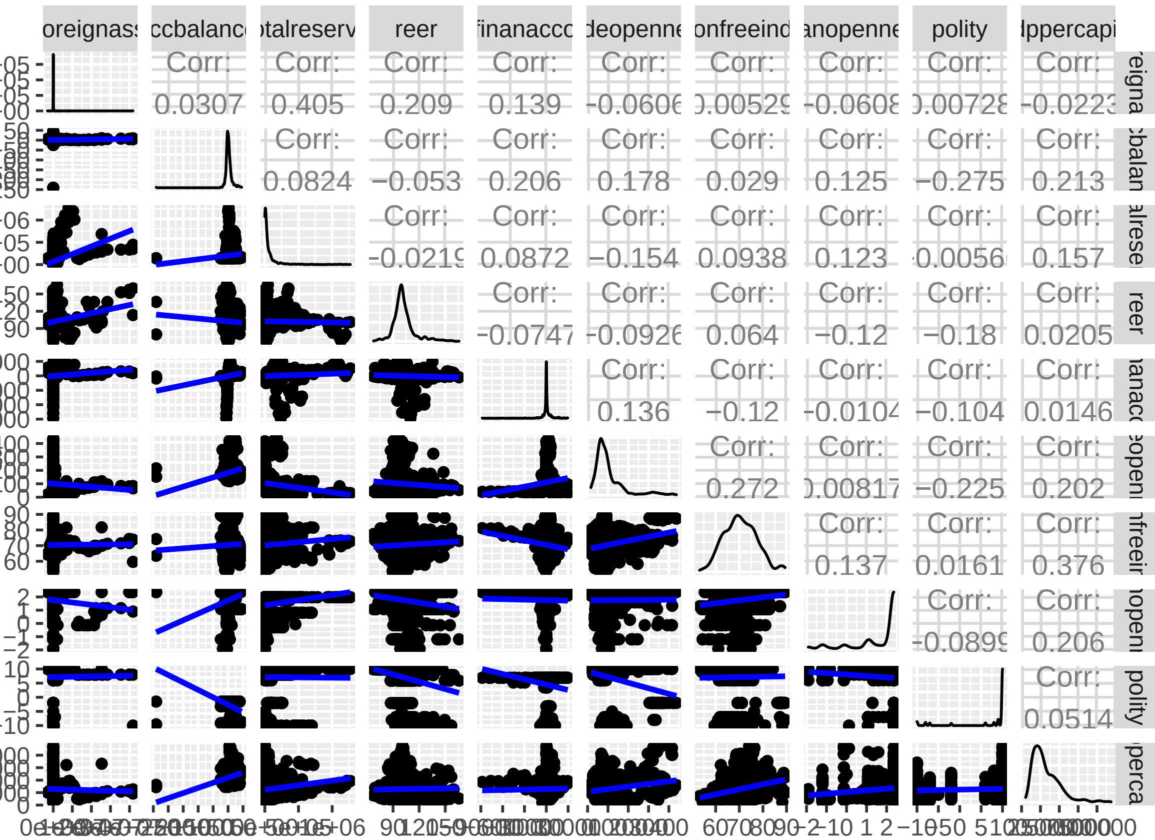
- **correlation heatmap**



From the color of the heatmap grids, we cannot spot many strong correlations. Polity score turns out to have a moderately negative correlation with current account balance (measured as % of GDP), suggesting that countries with relatively less stable rule of law or less robust democracy are tend to have larger current account balance as percent of their national income (which means that those countries run trade surplus). On the other hand, there seems to be a moderate, positive correlation between total reserve and net foreign assets. This makes sense, since both metrics measure the amount of a country's assets in foreign currency.

Next, the heatmap also reveals a positive relationship between economic freedom index and GDP per capita; freer economies were more likely to be wealthier. From the last quarter's EDA project, I discovered that it was a global trend reaching beyond high-income countries that richer countries were more likely to have more open, freer economies. What grabs our attention here is that such a relationship holds even if I limit the scope condition to high-income countries only. This relationship lends credibility to including the varieties of capitalism perpsective in the analysis. Liberal market economies, exemplified by the socio-economic structure of the United States, tend to have more open and freer economies. If they are more likely to have higher incomes, and if they are more integrated in international financial market and commercial economies (by being more "open" and "freer"), then they have both the capacity and willingness to maintain higher sovereign credit ratings.

- **bivariate scatterplots**

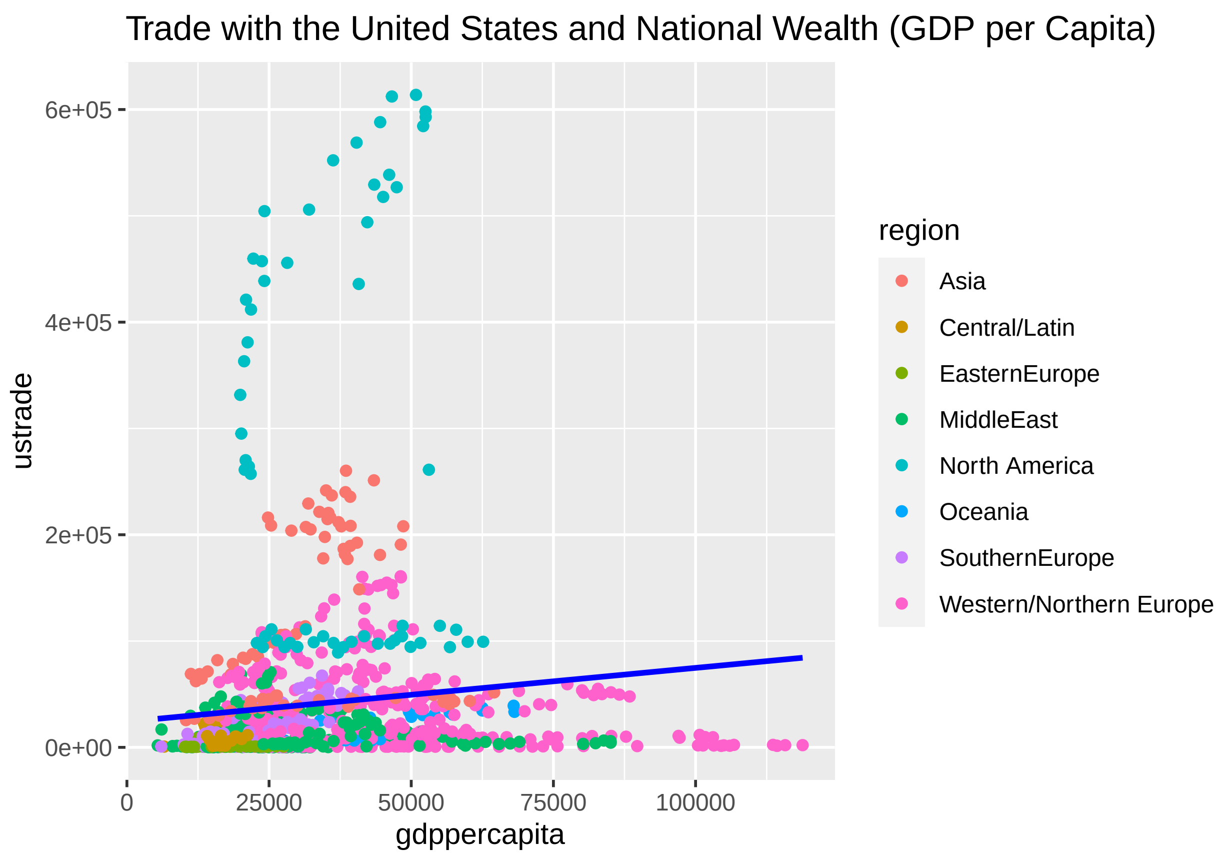


The pairwise scatterplots also support the findings from the previous heatmap plot. We do not see the evidence of high multicollinearity between variables here as well; rather, only moderate correlations are spotted.

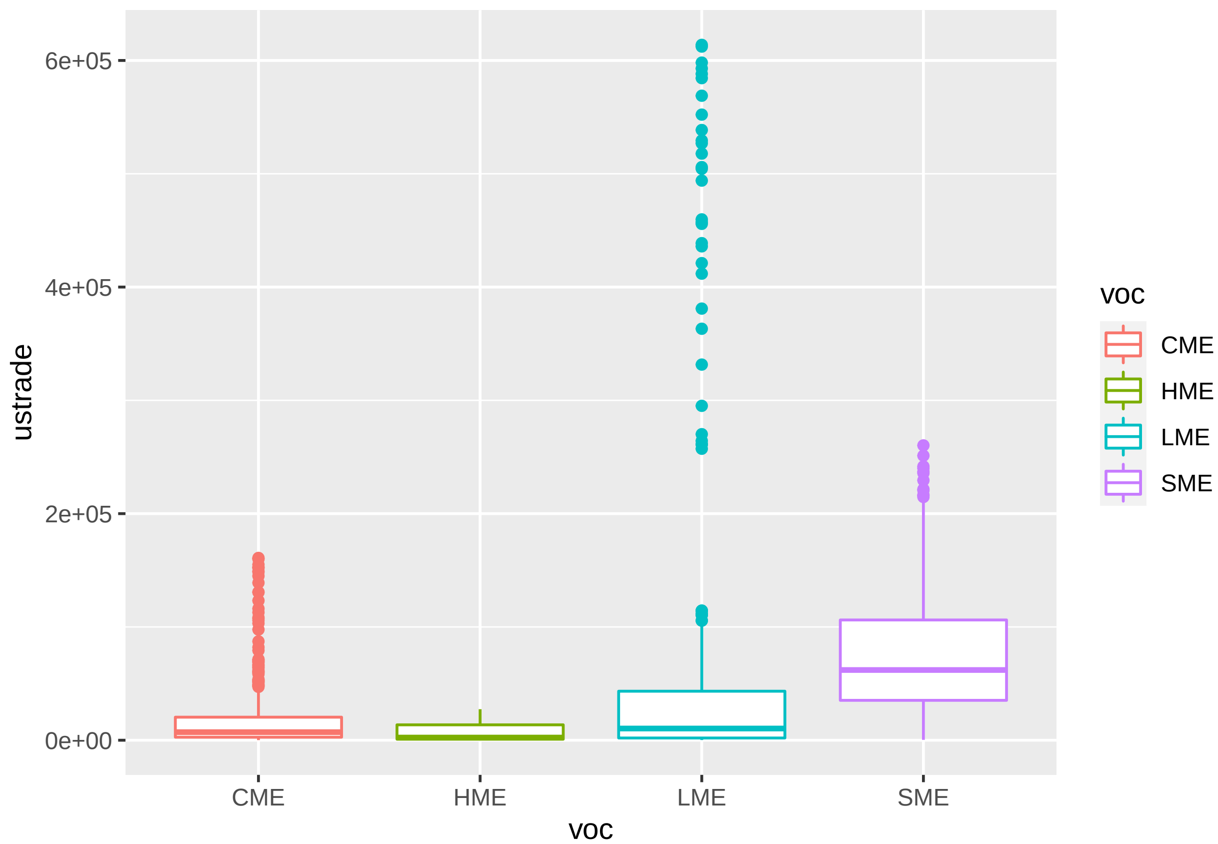
(3) Economic Ties with the United States

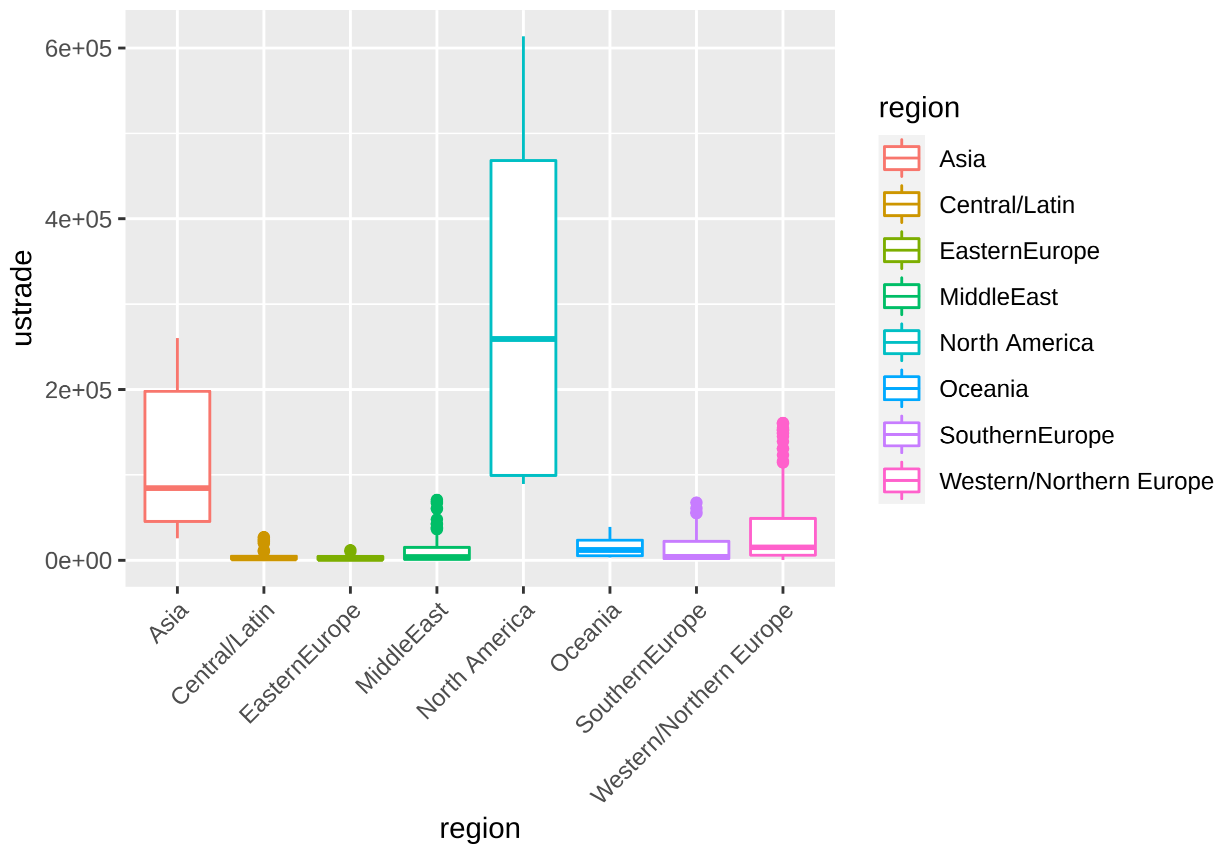
Lastly, I examine whether there is a pattern in the economic ties with the United States, notably in trade relationships. Contrary to Adam Smith's classic economic theory of comparative advantage and - according to it, only logical - international division division, economists instead have found, almost consistently, that there are more intra-industry trades and trades among high-income countries rather than trades based on comparative advantages/disadvantages. In the EDA project for the previous quarter, I also showed that there is a significant correlation between a country's income (measured by GDP per capita) and the volume of trade with the U.S., rather than a correlation associated with a country's comparative advantages based on relative factor endowment. I expand this analysis, and see whether high-income countries show different trade patterns with the United States based on their region and their institutional features: do certain regions trade more with the United States, and/or does the United States trade more with the countries which share institutional affinities with?

- **Trade Volume with the U.S.**



We can see a weak positive trend between the GDP per capita and trade volume (measured in US dollars) with the United States. Nevertheless, the trend is not very strong - in fact, Western/Northern European countries with the highest GDP per capita do not seem to trade a lot with the United States. On the other hand, among Asian countries, the correlation between a country's GDP level and trade with the U.S. appears to be fairly strong. We can see that among high-income countries, rather than the (per capita) GDP level, probably geographic proximity, economic patronage (as seen in Asian countries) or other factors might matter more in predicting trade volume with the U.S.


6e+05
4e+05
2e+05
0e+00
ustrade
CME
HME
LME
SME
voc
voc
CME
HME
LME
SME

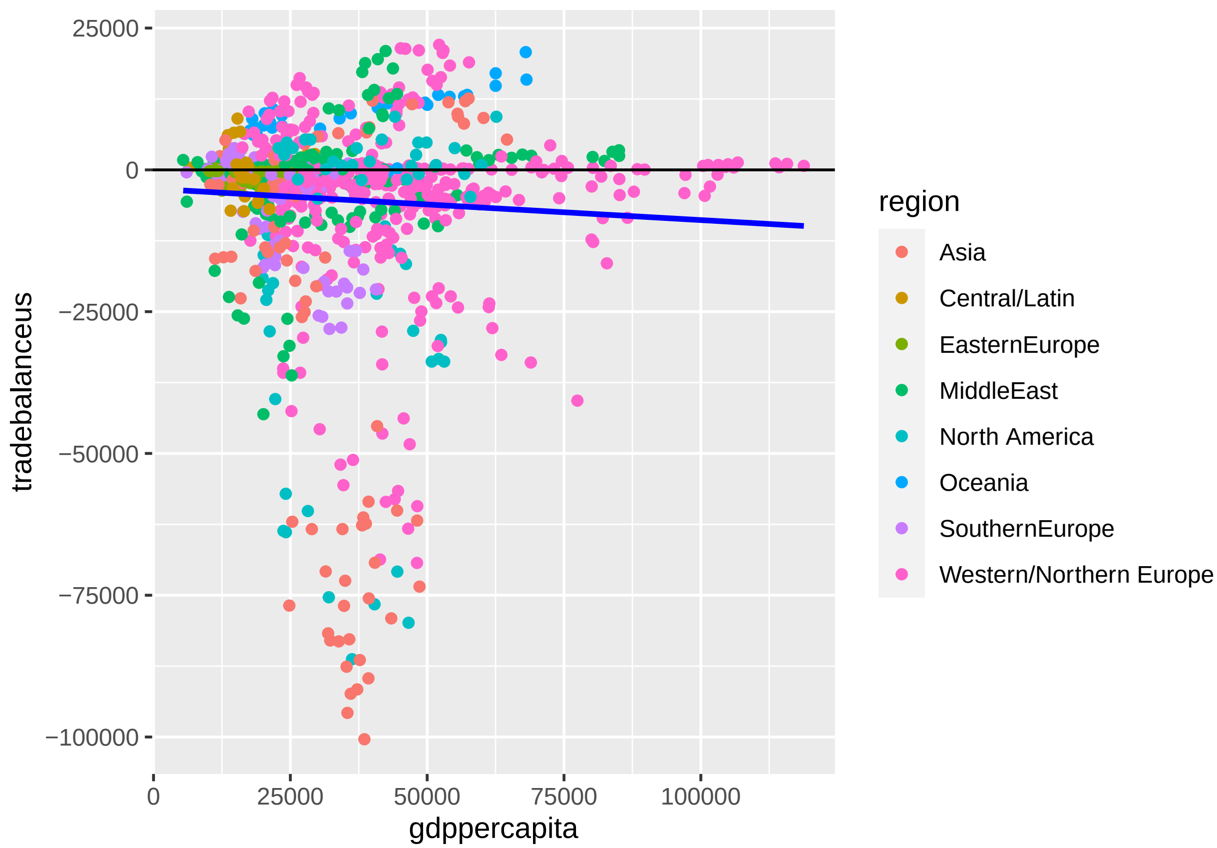


The boxplots above do support my aforementioned conjectures. The first boxplot shows the distribution of trade with the U.S. by varieties of capitalism. We can see that liberal market economies have a lot of "outliers" (to the extent that outliers themselves may have their own distribution!). Given that the United States is a quintessential example of the liberal market economic system, this supports the claim that those countries which share institutional and cultural (e.g., most liberal market economies are English-speaking, Anglo countries such as the United Kingdom or Australia) affinities with the U.S. trade more with it. Moreover, we can see that on average, state-centric market economies (SME's) trade more with the United States than the other two economies (CME's and HME's). SME's are composed of mostly Asian high-income countries, where the U.S. enjoys economic "patronage" over (e.g., South Korea and Japan). It is no wonder that indeed, these countries trade more with the United States.

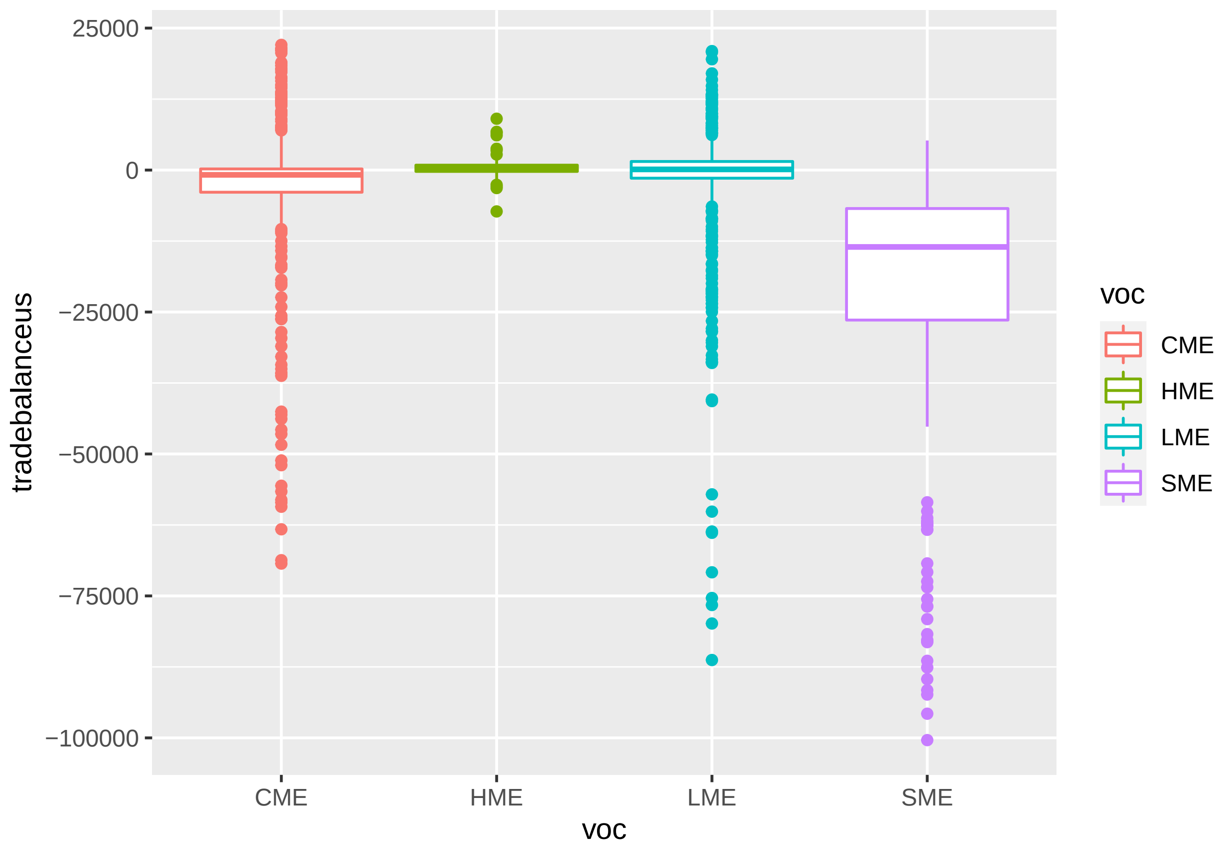
The second boxplot, on the other hand, supports the "gravitational theory of trade". Despite the fancy-sounding name, this theory basically argues that countries adjacent with each other trade more. Indeed, we can see that North America has the largest trade volume with the United States - *basically Canada* trade a lot with the U.S., followed by Asia and Western/Northern (i.e., industrialized) Europe.

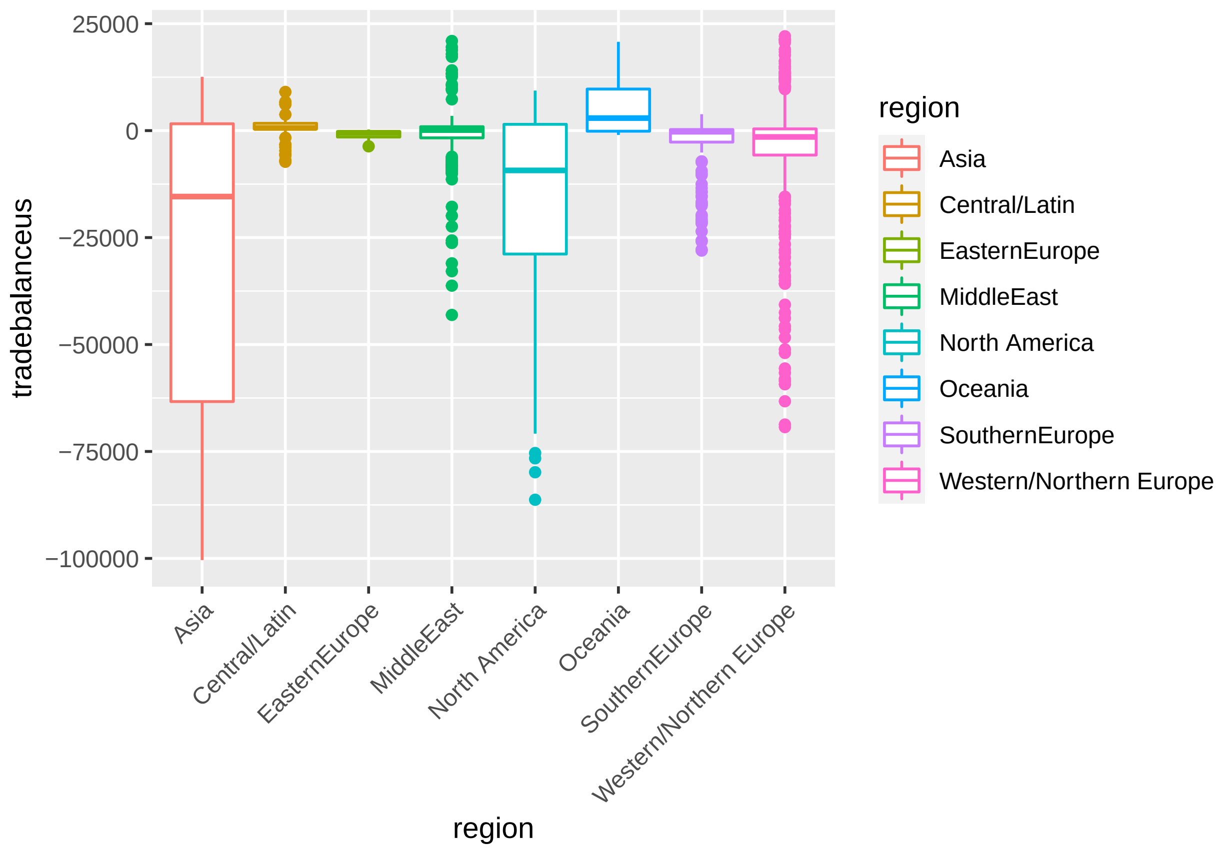
- **Trade Balance with the U.S.**

So far, we have examined high-income countries' economic ties with the United States measured by the magnitude of trade. Although trade volume is an important measure of economic relationship with a country, the contemporary American politics reveals that the American public also values "fairness" in trade - the rhetoric which President Trump emphasized for the 2016 election. This leads to my next question: then which countries have trade surplus with the United States? Is it also associated with a given country's wealth level?



From this graph, we can see that the U.S. tend to have trade deficits with the high-income countries - i.e., there are more observation points under the horizontal line of `tradebalanceus` $= 0$. Additionally, we can see that there is a very weak association between the GDP per capita and the trade balance of a country with the United States. The following boxplots show how the trade balance with the U.S. is distributed in each region and VoC.


25000
0
−25000
−50000
−75000
−100000
tradebalanceus
CME
HME
LME
SME
voc
voc
CME
HME
LME
SME



The U.S. runs the largest trade deficit in Asia on average, along wiht the Western/Northern European countries and with Canada. The fact that Asian countries are disproportionately state-centric market economies explains why the U.S. runs trade deficits with the SME's.

Model Building & Selection

The goal of this section is to identify the best model(s) associated with each method: OLS, fixed-effects models, and Lasso. For OLS and fixed-effects models, I specify the candidate models based on the existing literature and the results from the EDA above. On the other hand, since Lasso regression includes a penalty term in the model, I rely on its automated variable selection mechanism - that is, I try out the all-inclusive (i.e., "kitchen-sink") model and examine which coefficients are zeroed out. All in all, the motivating question for this section is as follows: which combination of predictors *predict* the sovereign credit ratings the best?

Second, to identify the best model(s) associated with OLS, fixed-effects and Lasso, I use k-fold cross validation (where k = 10). I select the models with the lowest test MSE, which will then advance to the final stage of this report (model validation).

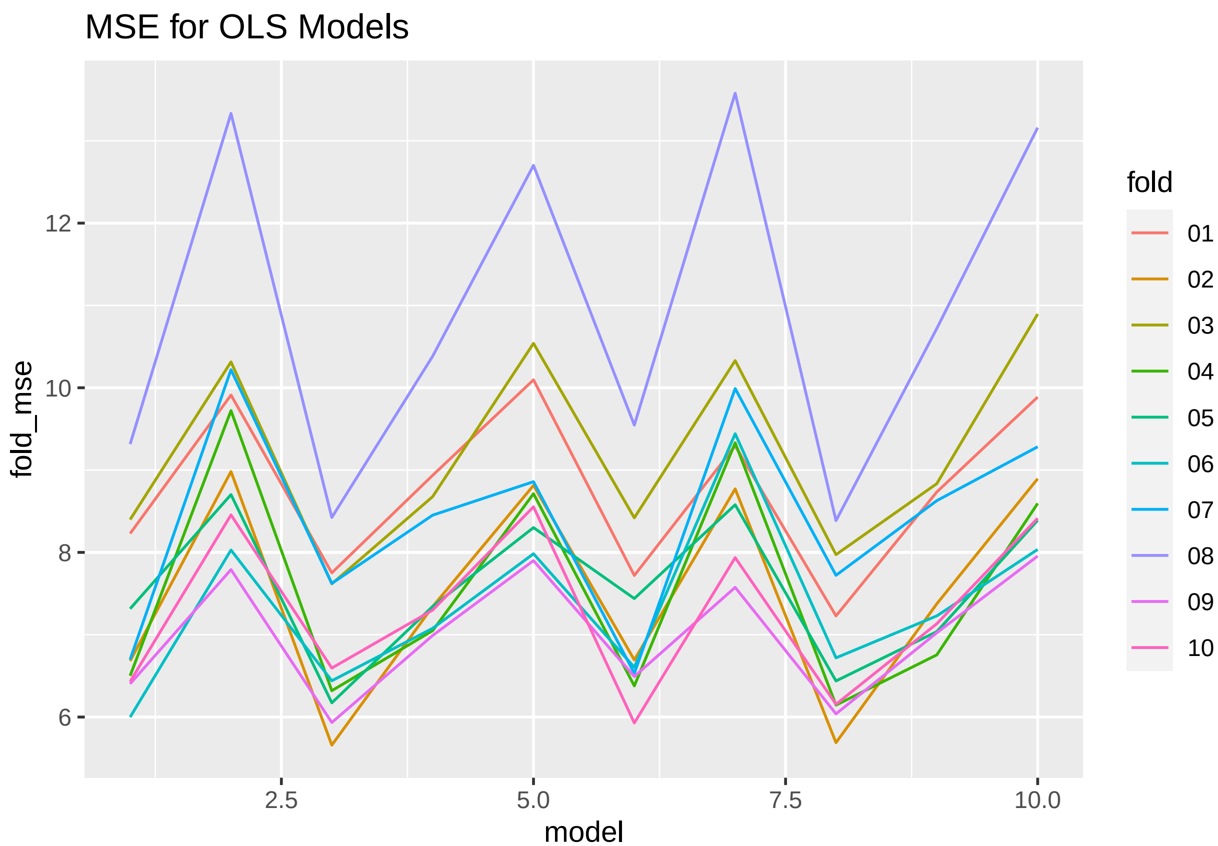
1. (Pooled) OLS Regression

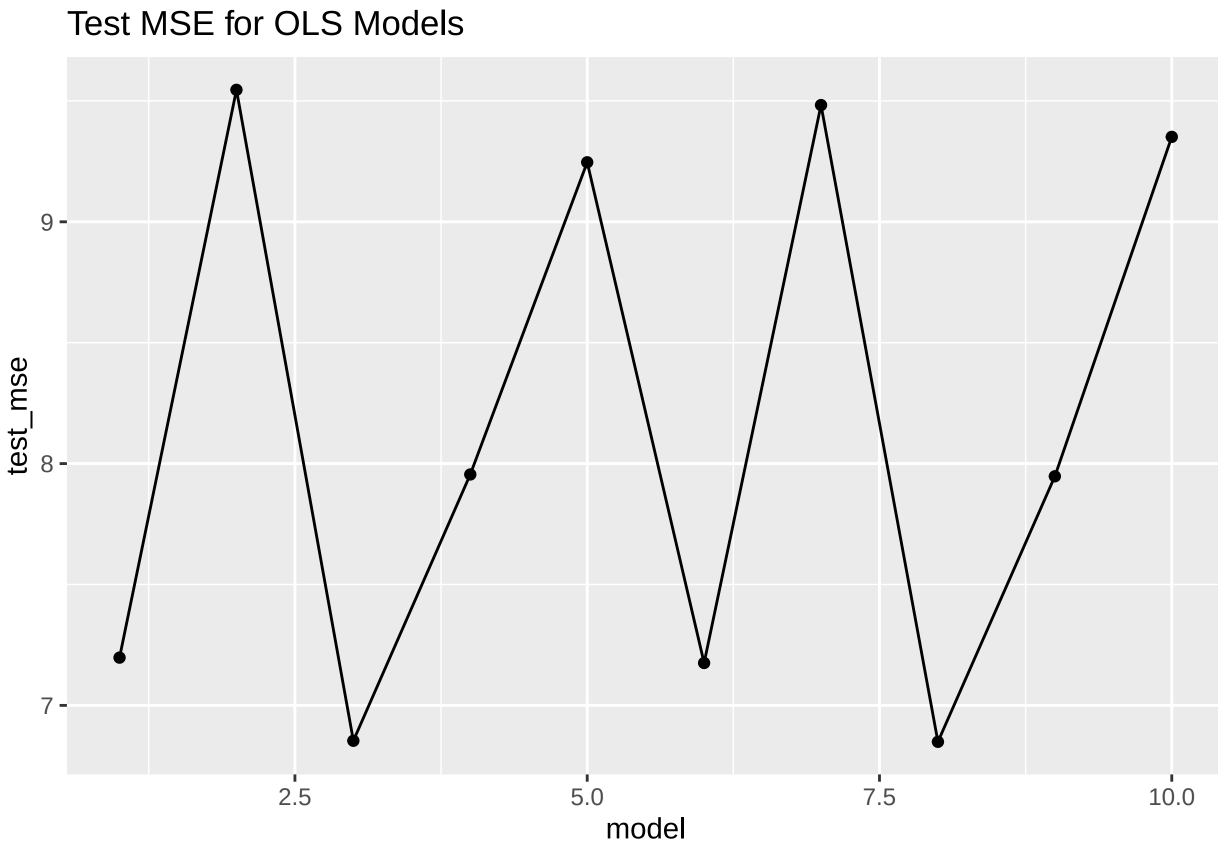
The first statistical method is the Ordinary Least Squares regressions. Using the training data, I fit the following five different groups of OLS models:

- (1) domestic fiscal model: sovereign ratings as a function of inflation, tax revenue, income and capital tax, government expense, government debt, GDP growth, and GDP per capita
- (2) international economic factors model: sovereign ratings as a function of net foreign assets, current account balance, total reserve, real effective interest rates, and net financial account

- (3) economic structure model: sovereign ratings as a function of trade openness, economic freedom index, and financial openness
- (4) domestic political economy model: sovereign ratings as a function of polity, inflation and unemployment
- (5) economic tie with the U.S. model: sovereign ratings as a function of trade wiht the U.S. and trade balance with the U.S.

In addition, models (6) to (10) are basically the same models but with one-year-lagged variables of the predictors added. I do not include 2-year or 3-year lagged terms because sovereign ratings change relatively quicly to reflect the domestic economic and political circumstances. Sovereign ratings sometimes fluctuates within a month, and even within the same week or day! Therefore, since economic or political "shocks" two or years ago are less likely to change the sovereign ratings of this year, I do not include further lagged terms because they are not theoretically relevant.






```
## # A tibble: 10 x 2
##    model test_mse
##    <int>    <dbl>
##  1     8     6.85
##  2     3     6.85
##  3     6     7.18
##  4     1     7.20
##  5     9     7.95
##  6     4     7.96
##  7     5     9.25
##  8    10     9.35
##  9     7     9.48
## 10     2     9.55
```

The k-fold cross-validation (k=10) results tell us that model (8) and model (3) do the best job in minimizing the test MSE (cf. since standardizing the data such that all the numerical variables' values have the mean of 0 and standard deviation of 1 does not change the "distribution" of the data per se, running the same models on the standardized data set does not change the selection of the candidate models). That is, economic structure model (with one-year lagged variables) does the best job in predicting the sovereign ratings. On the other hand, the worst models are international economic models and political models.

Given that a lot of lenders are not domestic but often international, the insight which we get from the pooled regressions (OLS) - that how much a country's economy is intergrated into the international economy seems to predict credit ratings - makes sense. As Axelrod and Keohane (1985)'s well-cited article implies, if a country's economy is more open to the international economy, the future costs it faces by defaulting on sovereign debts will be more costly. Therefore, (provided that they could) they choose not to default.

2. Taking Panel Structure into Account: Fixed-Effects Models

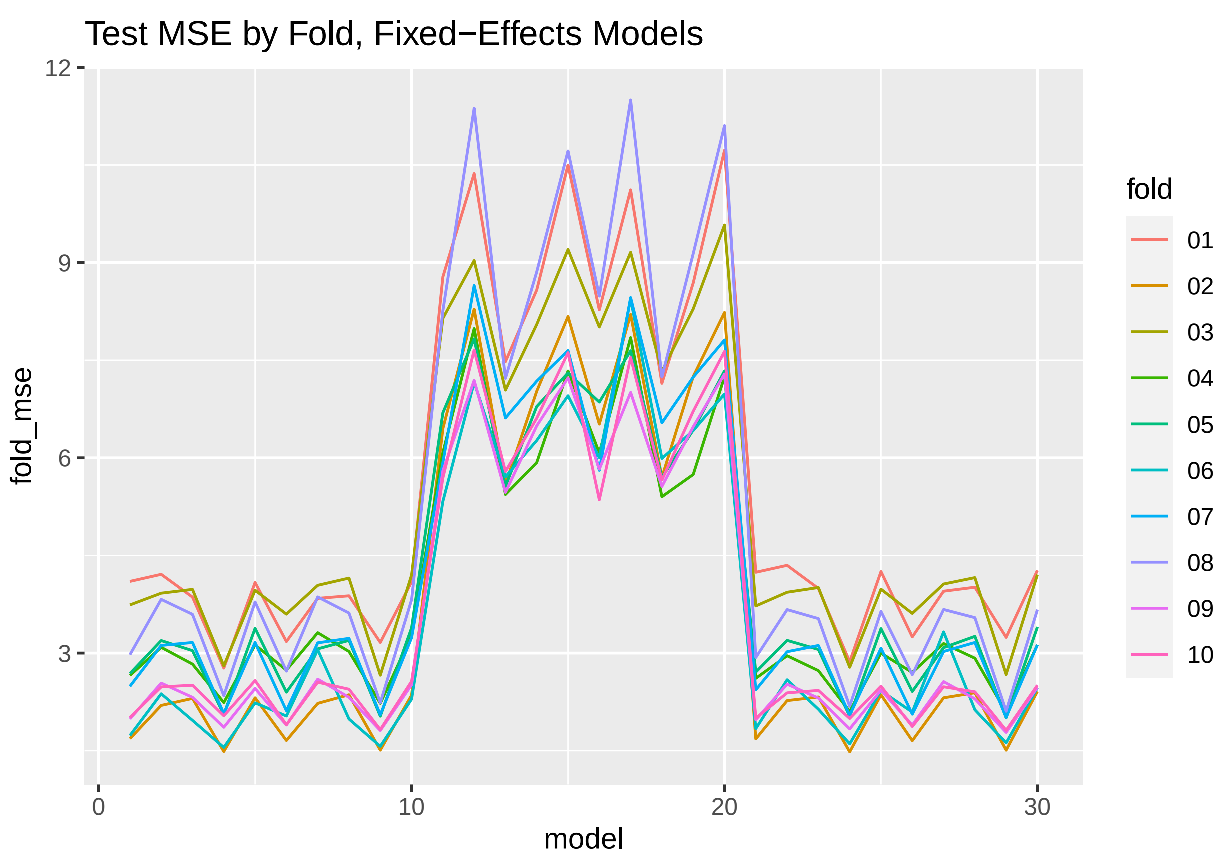
Although the previous results from OLS regressions suggests some interesting findings, note that the OLS models above are "pooled" models, where the country-specific differences are not accounted for in the model. Therefore, in this section, I account for the time-invariant, country-specific and institution-specific differences in the model.

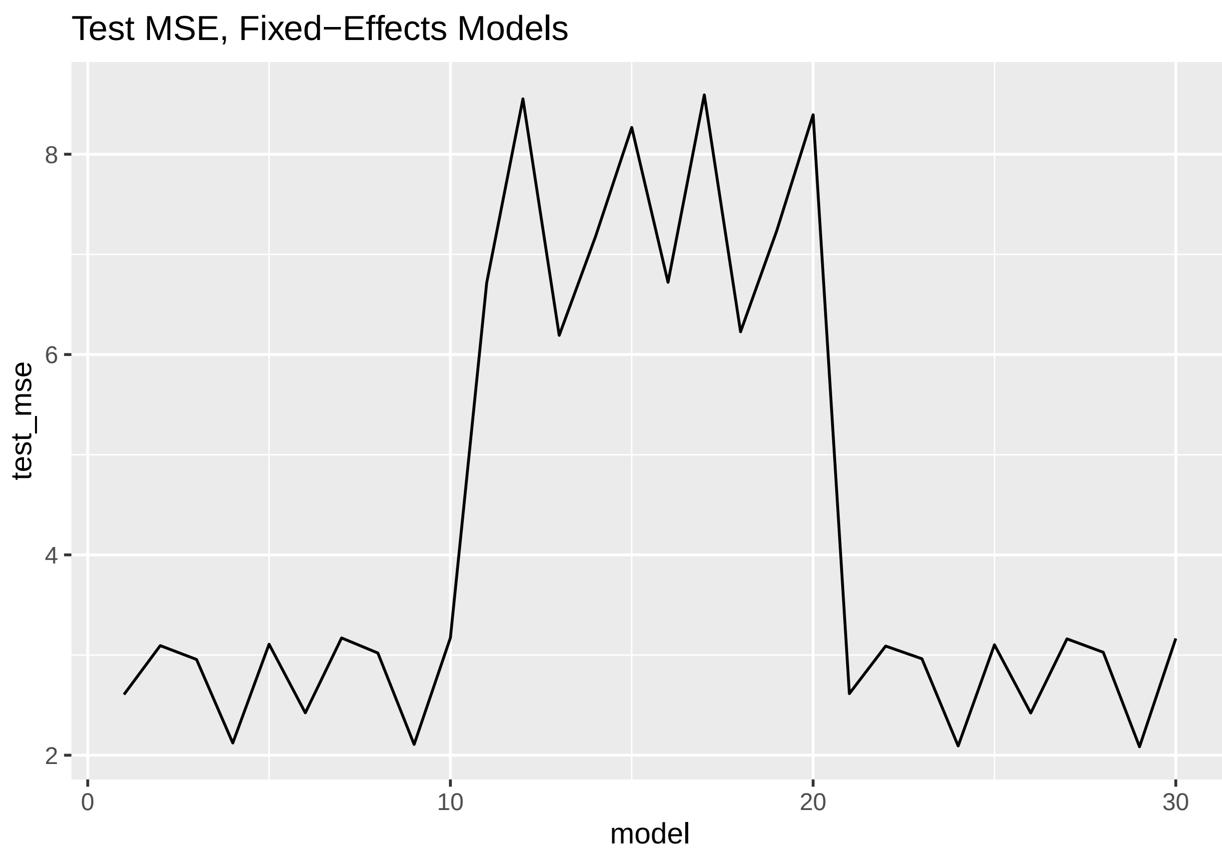
The `lmer()` function in `lme4` package allows us to fit mixed effects models - i.e., fitted a random-intercept model (with fixed slopes) to account for the time-invariant country-specific and voc-specific effects. Of course, there are more than one packages to allow estimating fixed and random effects models. I used `lmer()` function because it allows a more flexible estimation than other functions such as `plm()`.

Note that although I used the language "effects" (since this is the terminology), the primary purpose of this paper is not inferential; my purpose is not to identify the effects of different economic, political and social institutions on the sovereign ratings, but to account for the possibility that accounting for such possibilities (that there might be some country-level or institutional-level factors which might be associated with the ratings) would help increase the predictive power of the model.

Lastly, I choose to simply go with the "regular" standard errors. There are ways to correct for the violation of the core OLS assumption that error terms (hence observations) are identically and *independently* distributed (i.e., $N(0, \sigma^2)$). One such way is to use the panel-corrected standard errors to make viable inferences about the coefficient estimates. However, since I am mainly interested in prediction not inference, it would be okay to use the normal standard errors instead of the panel-corrected standard errors; hence, for this project, I go with the conventional standard errors (cf. I would have used panel-corrected SE if I were interested in identifying the *effects* of each predictor on ratings, though).

The models tested in this section are the same as the models I run in the OLS section: models (1) to (5) without lagged variables, and models (6) to (10) with one-year lagged variables. This time, I add country-fixed effects to models (1) to (10), varieties-of-capitalism (VoC)-fixed effects to models (11) to (20), and both country- and VoC-fixed effects to models (21) to (30). In total, I test 30 different models.






```
## # A tibble: 30 x 2
##    model test_mse
##    <int>    <dbl>
##  1    29     2.08
##  2    24     2.09
##  3     9     2.11
##  4     4     2.12
##  5    26     2.42
##  6     6     2.42
##  7     1     2.61
##  8    21     2.61
##  9     3     2.96
## 10    23     2.96
## # ... with 20 more rows
```

First, we can see that models with country-fixed effects did better in general. From the MSE's, we can tell that we should definitely include country-fixed effects in the model. On the other hand, controlling for more general, institutional/structural features of a country's economy (i.e., the "variety" of capitalism) did not contribute much to lowering test MSE. The second main takeaway is that controlling for the country-specific effects, domestic political economy models, where sovereign ratings are predicted as a function of polity, inflation and unemployment, did the best job lowering the test MSE in our 10-fold cross-validation.

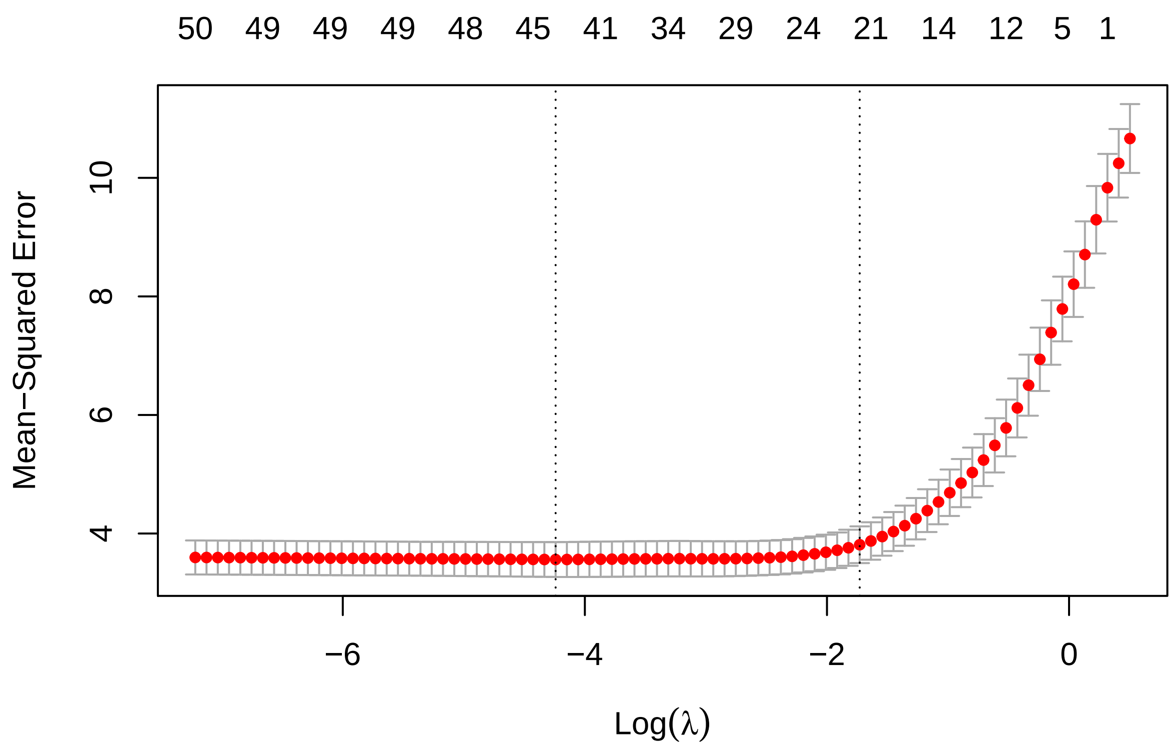
In sum, pooled OLS models with international economic integration factors (e.g., total reserve, net financial account, etc.) and country-fixed effects models with domestic political economic factors showed the best performance in minimizing the MSE. Therefore, these models are chosen as our "candidate" models, which are then to be proceeded to the model validation stage.

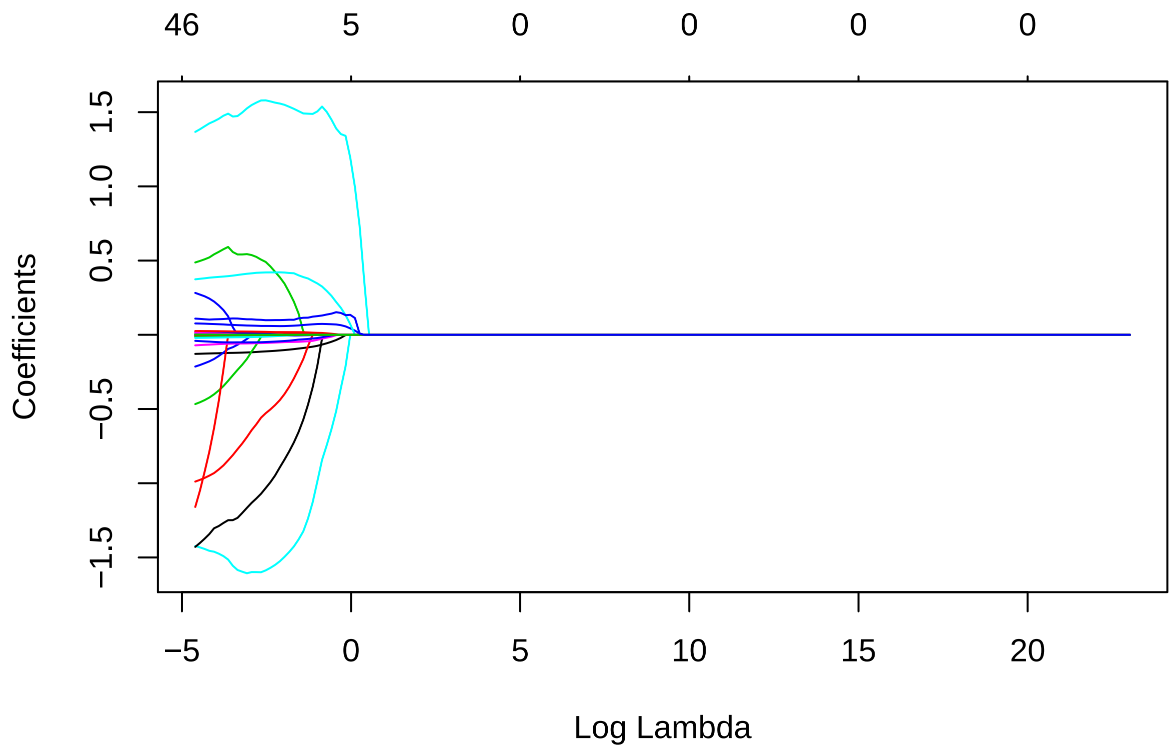
3. LASSO

In this section, the approach is somewhat different; it is not to try out different lasso regression models and choose the best model based on the cross-validation results (as above). Rather, I resort to the automated variable selection feature of Lasso regression. With the penalty term, it shrinks some of the β estimates to zero - I am interested mainly in (1) which variables' coefficients are zeroed out, (2) whether an automated selection model - lasso regression in this case - outperforms (i.e., does a better job) predicting the response variable than theoretically motivated models. Therefore, the LASSO model **does not** advance to the model-validation stage, because this is not theoretically interesting, nor part of the guiding research questions I aim to answer throughout this project. I also do not run the ridge regression, because the approach is somewhat similar (cf. note that we only have to change the code - to `alpha = 0` - slightly to run the ridge model).

The first plot: the results of cross-validation and how it "trains" the lambdas based on the MSE's. We can see that with the lambda that minimizes the cross-validation MSE - there are about 43 variables selected. A more regularizing lambda (to guard against the potential overfitting) zeroes out even more variables, shrinking the number of coefficients to about 21.

The next plot: shows how the coefficients were zeroed out - consistent with the previous plot, it shows that more and more coefficients are shrunken to zero as the value of (logged) lambda gets larger. Eventually, all variables are zeroed out.






```
## # A tibble: 1 x 2
##   lambda_min lambda_1se
##        <dbl>      <dbl>
## 1       3.81       4.65
```

With the training-error minimizing lambda, the test MSE is equal to 3.81344; with more regularizing lambda, the test MSE is 4.64653. Compared with the cross-validation test MSE from the OLS regressions, LASSO definitely did better (LASSO regression lowers MSE more than OLS regressions). However, when compared with fixed-effects models, LASSO actually did worse than some fixed-effects models! Impressively, taking into country-specific effects along with theory-based variables outperformed the mechanical, automated variable seletion algorithm.

Finally, I take a look at which variables LASSO zeroed out (and did not).

```
## # A tibble: 52 x 3
##    variable             min_lambda_coef lambda_1se_coef
##    <fct>                          <dbl>           <dbl>
##  1 (Intercept)                  263.            204.
##  2 year                          -0.126          -0.0976
##  3 vocHME                        -0.955          -0.312
##  4 vocLME                        -0.430          NA
##  5 vocSME                        -0.184          NA
##  6 regionCentral/Latin           -1.45           -1.44
##  7 regionEasternEurope           NA              NA
##  8 regionMiddleEast              -1.35           -0.743
##  9 regionNorth America           -0.841          NA
## 10 regionOceania                  0.516           0.242
## # ... with 42 more rows
```

With the "optimal lambda", which minimizes the test MSE for cross-validaton, we can see that the LASSO regression zeroed out a regional index variable (Eastern Europe), real effective exchange rates, GDP per capita growth, and two lagged variables (polity score and GDP per capita growth). With a more regularizing lambda, indeed more variables were zeroed out: some varieties-of-capitalism variables (LME and SME), regional index variables (Eastern Europe, Southern Europe and Middle East), tax revenue as % of GDP, government expense as % of GDP, polity score, trade balance with the U.S., net foreign assets, inflation, real effective exchange rates, GDP growth rate and GDP per capita growth, net financial account, trade openness, and most of the lagged variables (except government debt as % of GDP, trade with the U.S., current account balance as % of GDP, unemployment, real effective exchange rates and financial openness).

Model Validation

Using the 2016-2018 split, I now test the performance of the "select" models from OLS and fixed-effects models: which one is our true champion? And how good is her performance? After running the six candidate models chosen from the 10-fold cross-validation, I evaluate and rank the models using two different criteria: (1) MSE and (2) information criteria (AIC and BIC).

To recap, following are the models to be assessed on the validation set:

(1) Model 1: OLS (economic structure model)

(2) Model 2: OLS with lags (economic structure model)

(3) Model 3: country-fixed effects (domestic political economy model)

(4) Model 4: coutnry-fixed effects with one-year lags (domestic political economy model)

(5) Model 5: country-fixed effects + VoC-fixed effects (domestic political economy model)

(6) Model 6: country-fixed effects + VoC-fixed effects with one-year lags (domestic political economy model)

(1) Using MSE Criteria

```
## # A tibble: 6 x 2
##   model_name   mse
##   <chr>      <dbl>
## 1 mod_01      25.7
## 2 mod_02      25.8
## 3 mod_03      33.3
## 4 mod_05      33.3
## 5 mod_06      33.3
## 6 mod_04      33.3
```

Overall, the test MSE's on the validation set are generally higher than those from test errors from the cross-validation. This answers one of the questions for the model validation stage - i.e., how well do the models built on the previous time period (1989-2015) predict the sovereign ratings for the future time periods (2016-2018)?

Among the five models, we can see that models (1) and (2) - the OLS models - did not do bad. In fact, they outperformed the fixed-effects models in minimizing the MSE on the validation set!

(2) Using AIC/BIC Criteria

MSE provides an intuitive measure of model performance, since it quantifies "how much off" the predicted values are from the actually observed values. One downside of using MSE as our sole criteria is that as we add in more and more variables, we can reduce MSE even at the cost of overfitting. Therefore, I consider other criteria to assess the model fit. I use information criteria (AIC and BIC), since they penalize the number of parameters included in the model.

```
## # A tibble: 6 x 14
##   model_name   AIC   BIC data  model_fit r.squared adj.r.squared sigma statistic
```

```
##   <chr>      <dbl> <dbl> <lis> <list>        <dbl>         <dbl> <dbl>     <dbl>
## 1 mod_03      391.  408. <tib~ <lmerMod>    NA            NA     0.321      NA
## 2 mod_05      393.  413. <tib~ <lmerMod>    NA            NA     0.321      NA
## 3 mod_04      405.  432. <tib~ <lmerMod>    NA            NA     0.318      NA
## 4 mod_06      407.  437. <tib~ <lmerMod>    NA            NA     0.318      NA
## 5 mod_01      709.  723. <tib~ <lm>          0.537         0.527 2.82       53.6
## 6 mod_02      713.  737. <tib~ <lm>          0.542         0.521 2.84       26.8
## # ... with 5 more variables: p.value <dbl>, df <int>, logLik <dbl>,
## #   deviance <dbl>, df.residual <int>
```

In contrast with the MSE-based performance assessment, per both information criteria, the assessment of the model changes quite a bit! According to the information criteria, fixed effects models without lagged variables did the best, followed by fixed-effects models with lagged variables. On the other hand, OLS models did the worst. This is rather puzzling - information criteria take into account the number of variables included in the model, but model 1, 3 has similar number of variables (with 3 and 4 variables, respectively), while their AIC and BIC differ hugely. Log likelihood and deviance, which are both alternative measures of model fit, also suggest that fixed-effects model without lagged variables did considerably better than OLS models.

Conclusion

So far, we have explored how well a country's economic, political, structural and institutional features predict its sovereign ratings issued by U.S.-based rating agencies among high-income countries. Interestingly, the model validation results suggest somewhat contradictory conclusions about which model(s) is(are) truly the "best" model in predicting sovereign credit/debt ratings - according to the MSE criteria, it seems okay to go with the OLS models. On the other hand, per information criteria, it might be better to use country-fixed effects.

FURTHER STEPS

Lastly, I can definitely come up with viable "further steps" for future research. First, I limited the scope of anlaysis to high-income countries only for the sake of the project. However, the same analytical conventions laid out in this project can be applied beyond the high-income country groups. It would be instructive to compare the results between countries of different income groups. For example, if we find that different models do better for different income groups (i.e., for low-income countries, economic ties with the U.S. explain more variance in ratings than they do for high-income countries), those results can spur new line of thinking in the disciplines of economics and international political economy.

Second, we can include more variables, such as previous history of defaults. Based on the intellectual tradition of neoliberalism - namely Axelrod and Keohane (1985)'s classic article on the "shadow of the future" and the reputation costs, Michael Tomz argued that previous year's default impact current year's credit score for countries. Although I was not able to find the dataset on country-level defaults during 1989-2018, if I could obtain one, it will be interesting to examine *even controlling for the country's default history*, whether the social, economic, and political factors still well predict the ratings of the given year.

Third, one peculiar finding from the analysis was that the performance of the models on the validation set differed depending on which criteria I used - i.e., AIC/BIC vs. MSE. We can dig deeper into this, and identify what accounts for the differences.

Fourth, we can try out other dimension reduction algorithms (e.g., PCR? PLS?), and compare the results with those from Lasso regression. Perhaps we can come up with some ways to integrate dimension reduction approaches with regression method, which, to the best of my knowledge, has not been tried in International Political Economy. This will make methodological contributions to the substantive field.

Appendix

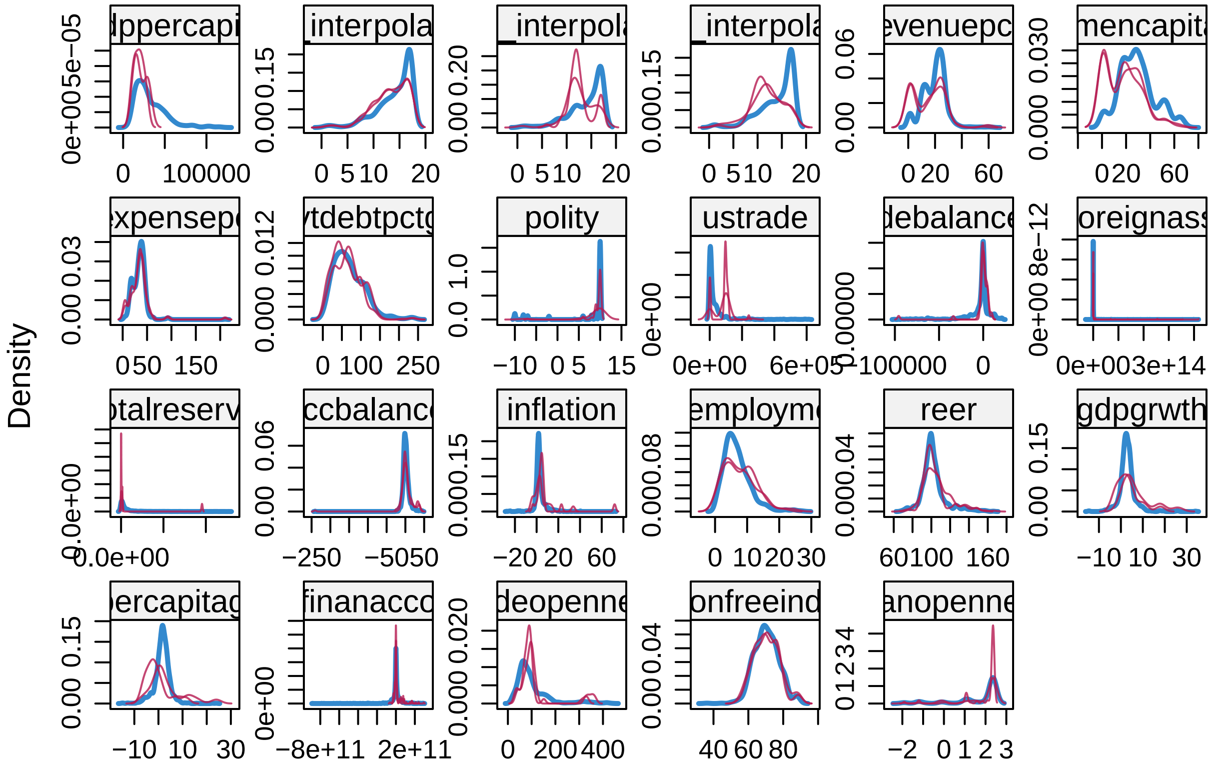
I. Codebook

No	S&P	Moody's	Fitch	Meaning
1	AAA	Aaa	AAA	Prime/Highest Quality
2	AA+	Aa1	AA+	High Quality
3	AA	Aa2	AA	High Quality
4	AA-	Aa3	AA-	High Quality
5	A+	A1	A+	Strong payment capacity
6	A	A2	A	Strong payment capacity
7	A-	A3	A-	Strong payment capacity
8	BBB+	Baa1	BBB+	Adequate payment capacity
9	BBB	Baa2	BBB	Adequate payment capacity
10	BBB-	Baa3	BBB-	Adequate payment capacity
11	BB+	Ba1	BB+	Likely to fulfill obligations, ongoing uncertainty
12	BB	Ba2	BB	Likely to fulfill obligations, ongoing uncertainty
13	BB-	Ba3	BB-	Likely to fulfill obligations, ongoing uncertainty
14	B+	B1	B+	High credit risk
15	B	B2	B	High credit risk
16	B-	B3	B-	High credit risk
17	CCC+	Caa1	CCC+	Very high credit risk
18	CCC	Caa2	CCC	Very high credit risk
19	CC	Caa	CC, C	Near default with possibility of recovery
20	SD, D	C	DDD, DD, D	Default

II. Post-Hoc Check of the Imputation Sets

Imputation Density Plots

The following density plots are the same one presented above in the EDA section. Note that the two red lines (imputed results) tend to overlap with each other.



III. Data Citation

1. Sovereign Rating Dataset

(1) Base dataset (1989-2006 ratings): from Professor Pedro Gomes at Birkbeck, University of London

(2) Expansion: S&P, Fitch, and Moody's website, cross-reference on https://tradingeconomics.com/ and https://countryeconomy.com/ratings/

2. Predictor Variables

Otherwise specified, predictor variables are collected from the World Bank Data portal.

(1) World Bank Data Portal: https://data.worldbank.org

(2) U.S. Bureau of Economic Analysis: https://www.bea.gov/data/economic-accounts/international

(3) Financial Openness Index: http://web.pdx.edu/~ito/Chinn-Ito_website.htm

(4) U.S. Census Bureau: : https://www.census.gov/foreign-trade/balance/index.html (dollar adjustment: https://stats.areppim.com/calc/calc_usdlrxdeflator.php)

(5) Polity Score: http://www.systemicpeace.org/inscrdata.html