Macalester College

Advanced Statistical Modeling

Class Project

Does Constitutional Form Affect a Nation's Income Level?

Author:
Kim Eng Ky

Advisor:
Prof. Christina Knudson

December 23, 2016

Abstract

Have you wondered why some countries have very different income levels? For example, what is the reason behind Cambodia being one of the poorest countries in the world and the United States is on the other end of the spectrum? In this paper, I model the response variable as levels of income classified by the World Bank which includes Low, Lower Middle, Upper Middle, and High Income. My main predictor of interest is constitutional form which includes Monarchy, Parliamentary, Presidential, and Socialist. I also consider other predictors such as cost of export, agricultural land, population density, whether or not a country is landlocked, and foreign direct investment. Only constitutional form and cost of export appear to be significant predictors of a nation's income level. I use multinomial logistic regression implemented in *nnet* package in R. More than half of the coefficients for the constitutional form variable are not statistically significant at 5% significance level so I am reluctant in concluding that constitutional form affects a nation's income level. Cost of export appears to be more statistically significant. Higher cost of export is associated with a higher probability of a country being in a low income group.

1 Introduction

Different countries can have very different levels of income. For example, Cambodia and the United States. Cambodia is one of the countries with the lowest income per capita in the world, whereas the United States is on the other end of the spectrum. If we look at all the countries around the world, we can see that income levels tend to very different for some countries. North American and European countries tend to have higher income per capita compared to Asian, African and South American countries. Figure 1 shows the map of all the countries color coded based on their income levels. From the map, we can see that different countries can have very different levels of income. What makes them so different?

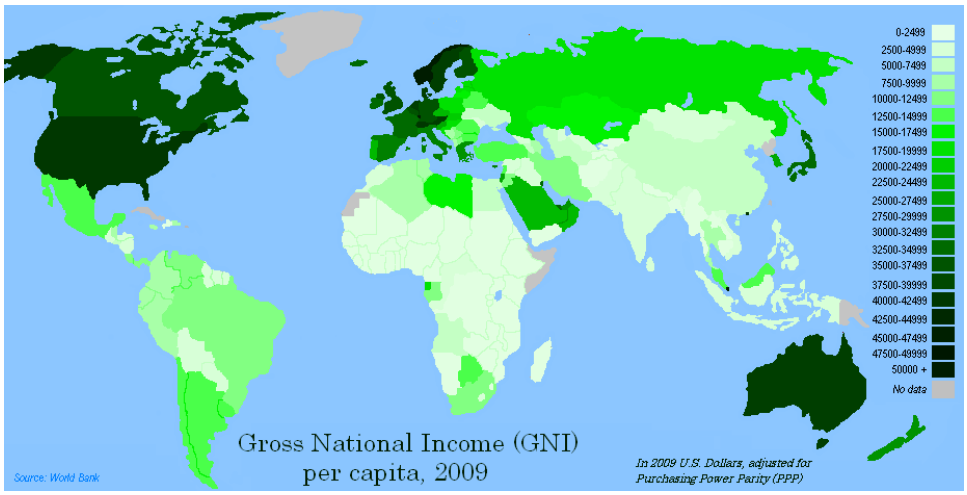

Figure 1: GNI per capita by Country
http://thosewhocansee.blogspot.com/2014/11/theres-something-about-teutonics.html

In this project, I would like to look at what determinants affect a country's income level. I am particularly interested in whether or not constitutional form affect the income level of the corresponding country. Upreti (2015) concludes that, for developing countries, a high volume of exports, plentiful of natural resources, longer life expectancy, and higher investment rates have positive impacts on a nation's income level. Birchfield and Crepaz (1998) look at the effects of constitutional structures on OECD (Organisation for Economic Co-operation and Development) countries on income inequality. They find that consensual constitutional structure is associated with lower income inequality. Though these studies answer different questions, they help me understand whether or not constitutional forms affect a nation's income level.

I hypothesize that constitutional form does affect a nation's welfare. A country with a

constitutional structure that is more democratic is likely associated with higher income level.

2 Data

I collect data from 171 countries in 2013. My response variable is the income levels obtained from the World Bank website (http://data.worldbank.org). This variable has four levels including Low Income (26 countries), Lower Middle Income (46 countries), Upper Middle Income (45 countries), and High Income (54 countries). The categories are based on the Gross National Income (GNI) per capita. GNI is measured as the income earned by a nation's residents both domestically and from abroad. A country with $1,045 or less GNI per capita is a low income country, between $1,046 and $4,125 is lower middle income, between $4,125 and $12,736 is upper middle income, and above $12,736 is high income.

In order to investigate what factors affect levels of income in each country, I look at variables such as constitutional form, population density, whether or not a country is landlocked, cost of export, agricultural land, and foreign direct investment. Though Upreti (2015) finds that life expectancy positively affects a nation's income level, I exclude this from my model because longer life expectancy is more likely the consequence of higher income than the cause of increasing in income.

I obtain the constitutional form variable by visiting the Wikipedia pages of every country in the world (http://en.wikipedia.org). Constitutional form is a four-level categorical variable which consists of monarchy, parliamentary, presidential, and socialist. A country is a monarchy if it is controlled solely by a monarch. For example, Oman is under the abosulte power of Sultan Qaboos bin Said al Said. A country is parliamentary if it is under a parliament or a parliament with a monarch where the monarch has little power. A parliament is a legislative, elected body of government. For example, Mauritius has a parliamentary system. Presidential constitutional form is a government body with president only, both president and prime minister, or mixed republican. For example, the United States is under a presidential constitution. The main difference between a parliamentary constitution and a presidential constitution is the way the leader, i.e. president or the prime minister, is elected. A socialist is a country where any of various economic and political theories advocating collective or governmental ownership and administration of the means of production and distribution of goods.

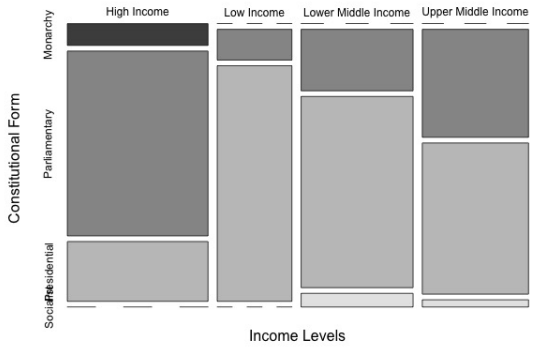

Figure 2: Constitutional Form

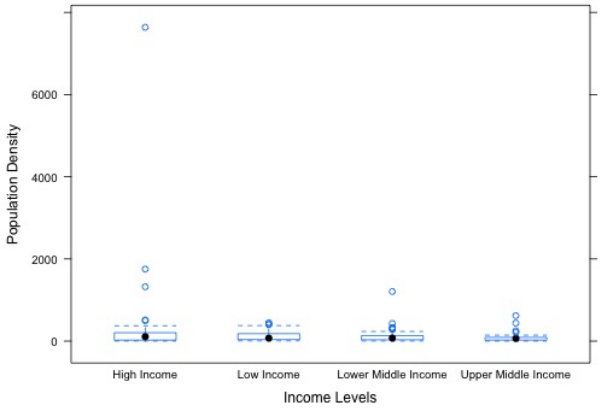

Figure 3: Population Density

Population density measures the number of people per one square mile. A country is landlocked if it does not border an ocean. Cost of export measures how much in U.S. dollars to export a 20-foot container. It includes legal fees, administrative fees, and the cost of shipping. Agricultural land tells what percentage of a country's land is agricultural. Foreign direct investment measures the net inflows (investments) from foreigners.

3 Exploratory Data Analysis

Before delving into modeling my data, I will do some exploratory data analysis. Figure 2 shows the plot of income levels versus constitutional forms. We can see that most countries are either presidential or parliamentary, with very few of them being monarchy and socialist. More than half of the high income group have parliamentary constitution, and more than half of the low, lower middle, and upper middle income groups are presidential.

Figure 3 shows the box plot of population density by income levels. There does not seem to be a significant difference between population density for these different income groups.
Figure 4 shows the plot of landlocked countries by income levels. There are more landlocked countries than those bordering with the ocean. About half of the countries in the low income group are landlocked where as a small percentage of high income group is landlocked.
Figure 5, 6, and 7 show the box plot of export cost, the percentage of agricultural land, and foreign direct investment , respectively, by income groups. There does not seem to have a significant difference in export cost, percentage of agricultural land, and foreign direct

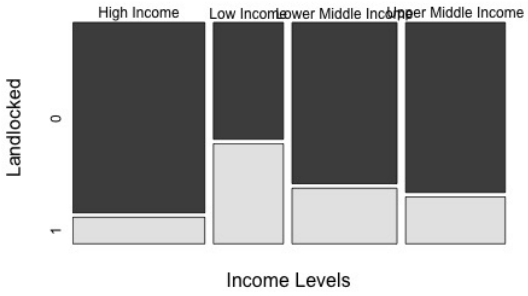

Figure 4: Landlocked

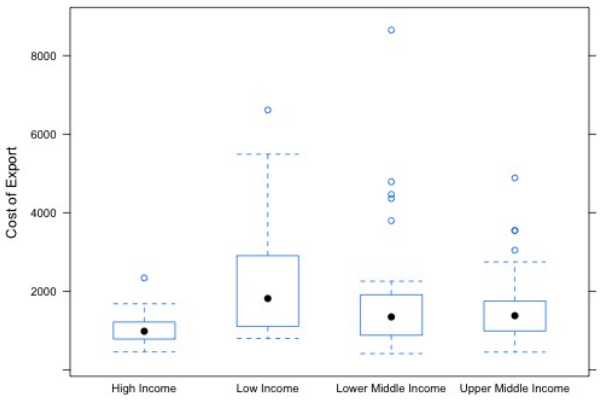

Figure 5: Cost of Export

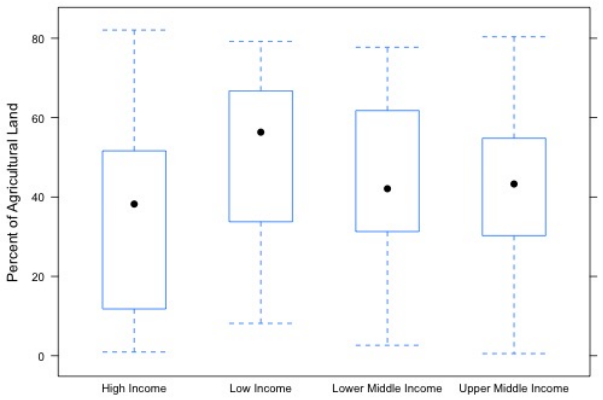

Figure 6: Agricultural Land

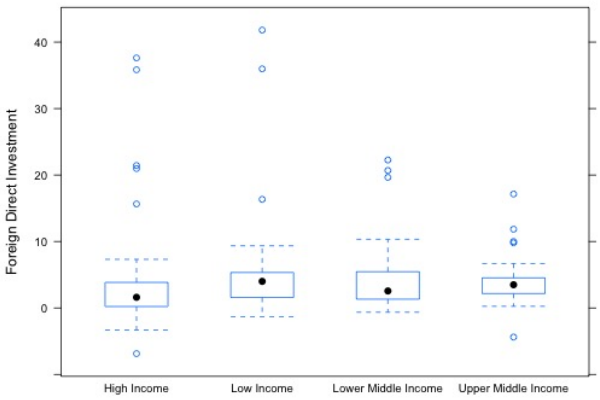

Figure 7: Foreign Direct Investment

investment between income groups.

4 Methodology

Since my response is a categorical variable, I cannot use linear model. And, because there are more than two levels in my response variable, I will use multinomial logistic regression to model my data. A multinomial logistic regression allows us to model a multi-level categorical response variable. We need to pick a baseline level, which can be any category of our response variable. In the case of four-level categorical response variable, we will have three different regression equations, each of which models the probabilities of the remaining categories compared to the baseline. The R package that I use to build my model is *nnet*. I mainly use the *multinom* command.

To decide which variables to include in my model as predictors, I use the forward selection method. With forward selection method, we keep adding one predictor at a time until having an additional predictor does not add significant goodness of fit to the model. Different criteria such as Akaike Information Criterion or log likelihood can be used. I use the likelihood ratio test (LRT), which tests if it is worth adding a predictor based on how it improves the log likelihood, with 5% significance level. With this method, I assume the simpler model, i.e. fewer predictors, is better.

I start by comparing the model with no predictors to models with one predictor. The LRT p-value for constitutional form is the smallest, and is lower than my cutoff at 0.05. Thus, my current model is a model with constitutional form as a sole predictor. I continuing adding one predictor to the model. The LRT p-value for cost of export is the smallest, and is lower than 0.05. Now my current model has constitutional form and cost of export as predictors. I continue adding more predictors but none show additional significant goodness of fit or likelihood to the model. I also look at the interaction between them, and it is not statistically significant. I log transform the cost of export because it is right-skewed.

5 Results

I choose high income group as the baseline for my multinomial logistic regression, so my regression equations can be written as:

$$\begin{aligned} log(\frac{\hat{p}_{low}}{\hat{p}_{high}}) &= -22.947 + 7.884\text{Parliamentary} + 10.600\text{Presidential} + -4.155\text{Socialist} \\ &\quad + 1.785\text{log(Export Cost)} \\ log(\frac{\hat{p}_{lowmid}}{\hat{p}_{high}}) &= 7.523 + 0.510\text{Parliamentary} + 2.296\text{Presidential} + 9.150\text{Socialist} \\ &\quad + 0.861\text{log(Export Cost)} \\ log(\frac{\hat{p}_{upmid}}{\hat{p}_{high}}) &= -16.897 + 9.632\text{Parliamentary} + 10.648\text{Presidential} + 17.111\text{Socialist} \\ &\quad + 0.965\text{log(Export Cost)} \end{aligned}$$

where p_{low} : probability of belonging to the low income group
p_{lowmid} : probability of belonging to the lower middle income group
p_{upmid} : probability of belonging to the upper middle group
p_{high} : probability of belonging to the high income group

Fixing cost of export at its mean value, table 1 shows the probability of each income level given a country's constitutional form. Note that 0's shown in the table do not suggest it is impossible, but the probabilities are infinitesimal which is essentially 0. Countries that are under a monarch are most likely a high income country with a probability of 80%. The probability of them belonging to the lower middle income group is 20%. They are not likely to be a low or upper middle income country at all. Countries with parliamentary constitution are most likely a high income country, and least likely a low income country, with probabilities of 45% and 4.5%, respectively. Countries under presidential constitution are most likely a lower middle income country with a probability of 36%. They are least likely a high income country with a probability of 14.5%. In this group, the probabilities are more spread out across different income groups compared to other constitutional categories. A socialist country is most likely a lower middle income country with a probability of 66%. A socialist is not at all likely to be a low or high income country.

Table 1: Probability of Each Income Level by Constitutional Form

	Monarchy	Parliamentary	Presidential	Socialist
p_{low}	0	0.045	0.216	0
p_{lowmid}	0.201	0.190	0.364	0.662
p_{upmid}	0	0.311	0.275	0.338
p_{high}	0.799	0.454	0.145	0

I check to see if fixing cost of export at a different value (its 25th and 75th percentile) would significantly affect the probabilities. The probability values slightly change, but it does not affect the ranking of probabilities for each income group. For example, fixing cost of export at its 25th percentile, countries with parliamentary constitution are still most likely a high income country, and least likely a low income country.

I then look at how cost of export affects the probabilities of each income level by fixing the constitutional form. I only look at countries with parliamentary or presidential constitutions because the other two, monarchy and socialist, contains too few observations. Figure 8 shows the odds of countries under parliamentary constitution belonging to a low, lower

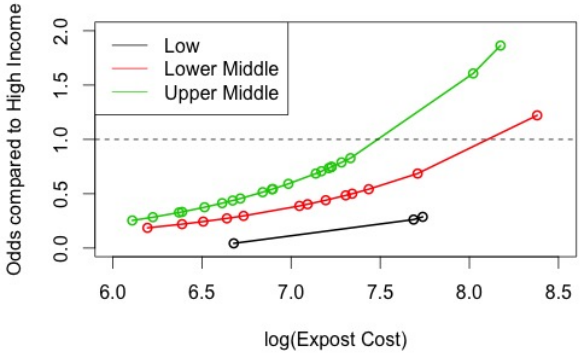

Figure 8: Parliamentary Only

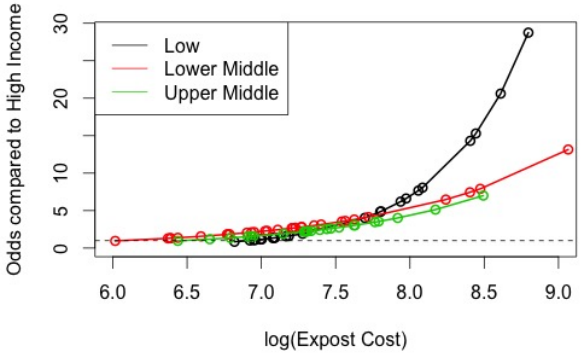

Figure 9: Presidential Only

middle and upper middle income group compared to a high income group. For all the three income groups, low, lower middle and upper middle, the odds compared to a high income group increase as cost of export increases. This suggests that it is less and less likely to be a high income country as the cost of export increases.

Figure 9 shows the odds of countries under presidential constitution belonging to a low, lower middle and upper middle income group compared to a high income group. Similarly, the odds of being a low, lower middle and upper middle income group compared to a high income group increases as cost of export increases. However, in this case, the odds increase much quicker than in the case of parliamentary constitution.

6 Discussion

From the results above, constitutional form and cost of export seem to explain a nation's income level. However, almost all of the coefficients in my model are statistically insignificant at 5% significance level using the Wald test (see appendix for the output of the regression). Table 2 shows the confidence intervals of all the coefficients found in my model. Values in bold mean they are statistically significantly different from 0 at 95% confidence level. As we can see, only five out of the 15 coefficients in total are statistically significant, i.e. different from 0, and none of which is consistently significant. Moreover, there are too few observations in the monarchy and socialist groups. For these reasons, I am reluctant in drawing inferences about the population from this model.

Table 2: 95% Confidence Intervals

	Low Income	Lower Middle Income	Upper Middle Income
Intercept	[-194.96, 149.068]	**[-14.40, -0.65]**	[-165.36, 131.57]
Parliamentary	[-163.97, 179.74]	[-1.75, 2.77]	[-138.70, 157.96]
Presidential	[-161.25, 182.45]	**[0.02, 4.57]**	[-137.68, 158.98]
Socialist	**[-4.16, -4.15]**	[-49.22, 67.52]	[-142.28, 176.50]
log(Export Cost)	**[0.71, 2.86]**	[-0.08, 1.81]	**[0.04, 1.89]**

Note: High income category is not in the table because it is the baseline, and all these confidence intervals are interpreted with respect to the high income group.

7 Conclusion

Because of the lack of significance in coefficients for the constitutional form variable, I would conclude that constitutional form might not be a significant predictor of a country's income level. The cost of export shows to be statistically significant at 5% significance level in two of the three models, and significant at 10% for the other one. Thus, the cost of export is a significant predictor of a country's income level.

This model could be improved by increasing sample size by having multiple years of data for each country, instead of a snapshot in time. Along with constitutional form as a predictor, political stability and economic institution may add significant improvement to the model. This then will lead to us implementing the model using generalized linear mixed model. A generalized linear mixed model allows us to investigate the fixed effects of the predictors on our response while taking into account the variability in the repeated measures of the same variables for the same countries.

References

[1] Birchfield, Vicki, and Markus ML Crepaz. "The impact of constitutional structures and collective and competitive veto points on income inequality in industrialized democracies." *European Journal of Political Research* 34.2 (1998): 175-200.

[2] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL *https://www.R-project.org/*.

[3] Upreti, Parash. "Factors Affecting Economic Growth in Developing Countries." (2015)

Appendix

```
> library(nnet)
> library(data.table)
> dat <- fread("income_level.csv")
> dat[, type2:=relevel(Type, ref="High Income")]
> currmod <-  multinom(type2~1, data = dat)
> mod1a <- multinom(type2~density,  data = dat)
> mod1b <- multinom(type2~Landlocked,  data = dat)
> mod1c <- multinom(type2~ConForm,  data = dat)
> mod1d <- multinom(type2~AgricLandPct,  data = dat)
> mod1e <- multinom(type2~ExportCost,  data = dat)
> mod1f <- multinom(type2~FDI,  data = dat)
> pchisq(2*as.numeric(logLik(mod1a)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.1902961
> pchisq(2*as.numeric(logLik(mod1b)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.006337534
> pchisq(2*as.numeric(logLik(mod1c)-logLik(currmod)), df=9, lower.tail = F)
[1] [1] 1.238767e-07
> pchisq(2*as.numeric(logLik(mod1d)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.004100498
> pchisq(2*as.numeric(logLik(mod1e)-logLik(mod0)), df=3, lower.tail = F)
[1] 3.243312e-06
> pchisq(2*as.numeric(logLik(mod1g)-logLik(mod0)), df=3, lower.tail = F)
[1] 0.6149674
> currmod <- mod1c #the lowest pval and is below 0.05
> mod2a <- multinom(type2~ConForm+density, data=dat)
> mod2b <- multinom(type2~ConForm+Landlocked, data=dat)
> mod2c <- multinom(type2~ConForm+AgricLandPct, data=dat)
> mod2d <- multinom(type2~ConForm+ExportCost, data=dat)
> mod2e <- multinom(type2~ConForm+FDI, data=dat)
> pchisq(2*as.numeric(logLik(mod2a)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.4845437
> pchisq(2*as.numeric(logLik(mod2b)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.005881791
> pchisq(2*as.numeric(logLik(mod2c)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.06095587
> pchisq(2*as.numeric(logLik(mod2d)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.002622122
> pchisq(2*as.numeric(logLik(mod2e)-logLik(currmod)), df=3, lower.tail = F)
[1] 0.6807999
> currmod <- mod2d
> mod3a <- multinom(type2~ConForm+ExportCost+density, data=dat)
```

```
    > mod3b <- multinom(type2~ConForm+ExportCost+AgricLandPct, data=dat)
    > mod3c <- multinom(type2~ConForm+ExportCost+Landlocked, data=dat)
    > mod3d <- multinom(type2~ConForm+ExportCost+FDI, data=dat)
    > pchisq(2*as.numeric(logLik(mod3a)-logLik(currmod)), df=3, lower.tail = F)
    [1] 0.6899981
    > pchisq(2*as.numeric(logLik(mod3b)-logLik(currmod)), df=3, lower.tail = F)
    [1] 0.1514192
    > pchisq(2*as.numeric(logLik(mod3c)-logLik(currmod)), df=3, lower.tail = F)
    [1] 0.227464
    > pchisq(2*as.numeric(logLik(mod3d)-logLik(currmod)), df=3, lower.tail = F)
    [1] 0.5926542
    > summary(currmod)
Call:
multinom(formula = type2 ~ ConForm + log(ExportCost), data = dat5)

Coefficients:
                   (Intercept) ConFormParliamentary ConFormPresidential
Low Income           -22.947438            7.8842031           10.599993
Lower Middle Income   -7.522796            0.5096967            2.296484
Upper Middle Income  -16.896980            9.6318753           10.647513
                   ConFormSocialist log(ExportCost)
Low Income                -4.155053       1.7853552
Lower Middle Income        9.149959       0.8607348
Upper Middle Income       17.110657       0.9648736

Std. Errors:
                   (Intercept) ConFormParliamentary ConFormPresidential
Low Income           87.764454            87.681825           87.680567
Lower Middle Income   3.507617             1.151861            1.161629
Upper Middle Income  75.750091            75.679277           75.679689
                   ConFormSocialist log(ExportCost)
Low Income             8.807119e-04       0.5467971
Lower Middle Income    2.977946e+01       0.4821119
Upper Middle Income    8.132274e+01       0.4743311

Residual Deviance: 401.6261
AIC: 431.6261
```