Assignmet 2

Dipa Rai

3/28/2021

Introduction

Democratization in many ways such as through technical assistance focusing on election processes, promoting civil society organizations, improving education and economy in the recipient countries. Aid can enhance human development by financing public expenditures such as quality education, health, infrastructure, and so on. This may ultimately lead to democracy as people become better educated and more well-off. Because educated voters are an important prerequisite for a well-functioning democracy (Gomanee,et.al.2005). Aid can promote democracy directly when it funds institutional development within the government and civil society and empowering oppressed groups. Initially, Seymour M. Lipset argued that the economic growth lead to democracy (Lipset, 1959). By promoting economic and social development, aid may provide the key inputs that could ultimately push a recipient country towards democracy.

This paper will try to look at the main relationship between aid and democracy using multivariate graphs. Potentially this analysis will use democracy as a binary dependent variable and try to predict whether the a county falls into democracy or into non-democracy categories. For this analysis Linear probability and tree classification models will be used.

Motivation

Erasmus Kersting and Christopher Kilby argue that aid can play potential roles in the promotion of democracy. Over a long period of time, aid may promote social and economic developments, which are considered to be prerequisites for democracy. In a short period of time aid may provide essential key inputs to elections. It is in the sense that in the election some aid could target opposition parties to assists pro-democratic parties and provide technical assistance for legislative reforms supporting democratic parties. Also aid might act as an incentive for democratic reform that aid can be provided to recipient in condition to be democratic (Kersting and Kilby, 2014) However, Knock in his multivariate analysis on the impact of aid on democratization over the 1975-2000 period found no such evidence that aid promotes democracy (Knack, 2004).

Data sets and Methods used

The primary data set is (aid) data set from Erasmus Kersting and Christopher Kilby (2014) from their analysis of aid and democracy, which is a country-level data. It contains key information on democracy indicators such as aid flows and economic variables. Also, in this analysis other data sets are used by merging which contained aid and democracy related

variables. Data sets such as (Polity IV) which contains political indicators such as political regime characteristics and transitions, (MEPV) data set contains Political violence and Conflict by regions, (sfi) contains state fragility matrix. The additional data sets are being filtered by year 2012.

primarily democracy id measured based on the indicators provided in the polity IV data set. The data set provides a 10 point scale of democracy based on evaluations of how executive are selected, and whether or not there are effective institutions checks and balances on executives power (Knack, 2004).

All the relevant libraries for this project

```
    library(rio)
    library(ggplot2)
    library(knitr)
    library(pander)
    library(car)
```

```
## Loading required package: carData
```

```
    library(stargazer)
```

```
##
## Please cite as:
```

```
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary
Statistics Tables.
```

```
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
    library(MASS)
    library(leaps)
    library(rpart)
    library(rpart.plot)
    library(randomForest)
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
    library(pROC)
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var

    library(dplyr)

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:randomForest':
##
##     combine

## The following object is masked from 'package:MASS':
##
##     select

## The following object is masked from 'package:car':
##
##     recode

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

aid <- import("http://www.joselkink.net/files/data/KerstingKilby2014-Tab-2-5.dta")
```

Reshaping data

Merge data file that translates country names in aid data to those in the Polity IV project

```
matchNames <- import("http://www.joselkink.net/files/data/KerstingKilby-Polity-name-matching.xlsx")

# Remove Ivory Coast due to problems with spelling on Windows computers
matchNames <- matchNames[-35, ]
aid <- aid[-35, ]

aid <- merge(aid, matchNames, by.x = "name", by.y = "kk", all = TRUE)
```

Mearging data sets

```
# Merge the Polity IV democracy data by Systemic Peace
p4 <- subset(import("http://www.systemicpeace.org/inscr/p4v2016.sav"), year == 2012)
names(p4) <- paste0("P4_", names(p4))
aid <- merge(aid, p4, by.x = "p4", by.y = "P4_country", all = TRUE)
```

```
# Merge the Major Episodes of Political Violence data by Systemic Peace
mepv <- subset(import("http://www.systemicpeace.org/inscr/MEPVv2016.sav"),
year == 2012)
names(mepv) <- paste0("MEPV_", names(mepv))
aid <- merge(aid, mepv, by.x = "p4", by.y = "MEPV_country", all = TRUE)

# Merge the State Fragility Index data by Systemic Peace
sfi <- subset(import("http://www.systemicpeace.org/inscr/SFIv2016.sav"), year
== 2012)
names(sfi) <- paste0("SFI_", names(sfi))
aid <- merge(aid, sfi, by.x = "p4", by.y = "SFI_country", all = TRUE)
```

Computed variables

The data set only contains original Freedom House scores and their changes, so it is computed to get the end result below in the first chunk. In the second chunk, a dummy variable for democracy based on the two measures of the Polity IV and Freedom house is created. in the third chunk a dummy variables for the initial level of democracy and initial FHdemocracy are recoded.

```
aid <- aid %>% mutate(
  CL_score = Initial_CL_score + .5 * (DCL_score_LL + DCL_score_UL),
  FH_score = Initial_FH_score + .5 * (DFH_score_LL + DFH_score_UL),
  PR_score = Initial_PR_score + .5 * (DPR_score_LL + DPR_score_UL),

  #Democracy Dummy
  democracy = as.factor(ifelse(P4_polity2 > 5, "Democracy", "Non-
Democracy")),
  FHdemocracy = as.factor(ifelse(CL_score > 30 & PR_score > 20,
                                 "Democracy", "Non-Democracy")),

  # A dummy variables for the initial level of democracy
  initial_democracy = car::recode(Initial_Polity, "-10:5=0; 6:10=1; else=NA",
                                  as.factor = FALSE),
  initial_FHdemocracy = ifelse(Initial_CL_score > 30 & Initial_PR_score > 20,
1, 0),

  # logged versions of the two aid variables,
  # Aid per Capita and Aid over GDP.
  logAPC = log(Aid_per_capita),
  logDAC = log(DAC_aid)
)
```

Nominal level variable for colonial empire that a country belonged to coded as 1.

```
aid$colonialPower <- "Independent"
aid$colonialPower[aid$British == 1] <- "British"
aid$colonialPower[aid$French == 1] <- "French"
aid$colonialPower[aid$Spanish == 1] <- "Spanish"
aid$colonialPower[aid$Portugese == 1] <- "Portugese"
```

```
aid$colonialPower[aid$Belgian == 1] <- "Dutch"
aid$colonialPower[aid$Belgian == 1] <- "Belgian"
aid$colonialPower[aid$Italian == 1] <- "Italian"
aid$colonialPower[aid$German == 1] <- "German"
aid$colonialPower[aid$American == 1] <- "American"
aid$colonialPower <- as.factor(aid$colonialPower)
```

Colonial empire variable, reduced the number of categories

```
aid$mainColonialPower <- car::recode(aid$colonialPower,
  "c('American', 'Belgian', 'German', 'Italian', 'Portugese')='Other'")
```

Nominal level variable for the region the country is in

```
aid$region <- ""
aid$region[aid$LAC == 1] <- "Latin America / Caribbean"
aid$region[aid$MENA == 1] <- "Middle East / North Africa"
aid$region[aid$SSA == 1] <- "Sub-Saharan Africa"
aid$region[aid$ECA == 1] <- "Europe / Central Asia"
aid$region[aid$SA == 1] <- "South Asia"
aid$region[aid$region == ""] <- NA
aid$region <- as.factor(aid$region)
```

Dependent dummy variable

The dependent variable here is democracy and it has to be a categorical variable to predict the probability of falling whether into autocracy or democracy categories based on the independent variable aid per capita. Since it is a supervised machine learning, we know that from the data set, the Polity IV that democracy is a continuous variable. Where it is measured in 0 to 10 scale where 0 being the most autocratic and 10 being the most democratic countries.

To make democracy a categorical variable, here a binary democracyDummy variable is created where there are only two categories one and zero. The one represent Democracy and the zero represent Non-Democracy. For the further analysis this binary democracyDummy will be used as a dependent variable.

```
aid <- aid %>% mutate(
  democracyDummy = na_if(recode(FHdemocracy,
                                'Democracy' = 1, 'Non-Democracy' = 0,
                                .default = -1), -1)
)
```

¸Cross table of the binary dependent variable

```
pander(table(aid$democracyDummy, aid$FHdemocracy, exclude = NULL))
```

	Democracy	Non-Democracy	NA
0	0	61	0
1	60	0	0

NA 0 0 83

The cross table above shows the categories of the democracy dummy variables. There are 60 countries which fall into category 1, which is democratic and 61 countries fall into 0 which is non-democracy and NA means missing data. The threshold number for democracy and non-democracy had already been designed above based on Polity IV scores of democracy which ranged from 0 to 10 scale.

Multiple regression to estimate positive relationship etween aid and democracy

```
stargazer(m1 <- lm(democracyDummy ~ Aid_per_capita, data = aid),
          m2 <- lm(democracyDummy ~ Aid_per_capita +
                     Initial_GDP_per_capita + Initial_population +
                     Ethnic_Fractionalization + initial_FHdemocracy, data =
aid),
          type = "text")
```

```
===============================================================================
= Dependent variable:
——————————————————————————————- democracyDummy
(1) (2)
————————————————————————————————————– Aid_per_capita 0.003*** 0.002**
(0.001) (0.001)

Initial_GDP_per_capita 0.018
(0.044)

Initial_population 0.031
(0.028)

Ethnic_Fractionalization -0.422**
(0.174)

initial_FHdemocracy 0.245**
(0.105)

Constant 0.359*** -0.097
(0.059) (0.638)

-------------------------------------------------------------------------------

Observations 121 120
R2 0.090 0.206
Adjusted R2 0.082 0.171
Residual Std. Error 0.481 (df = 119) 0.457 (df = 114)
F Statistic 11.788*** (df = 1; 119) 5.918*** (df = 5; 114)
===============================================================================
= Note: p<0.1; p<0.05; p<0.01
```

```
designMatrix <- m2$model

olsPredicted <- ifelse(predict(m2) > 0.5, "Democracy", "Non democracy")
```

In this multiple regression analysis the impact of aid on democracy is examined. The result shows that aid has significant positive impact on democracy. Whereas, the initial GDP per capipa does not have much impact on democracy. Thus, according to this result it could be said that aid can promote democracy.

Linear probability model

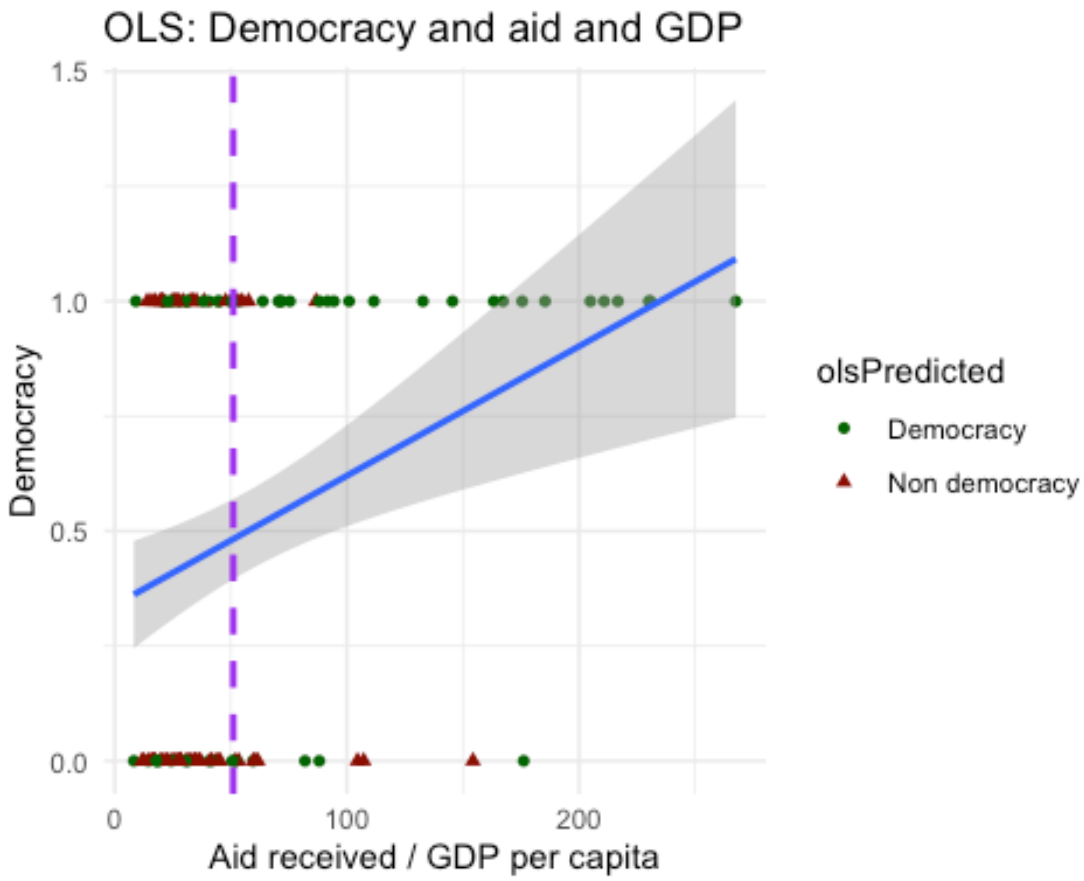
Classification regression and multivariate graph with a decision boundary

```
b1 <- coef(m1)

ggplot(designMatrix, aes(x = Aid_per_capita + Initial_GDP_per_capita , y =
democracyDummy)) +
  geom_point(aes(shape = olsPredicted, color = olsPredicted), height = .1) +
  geom_smooth(method = "lm", se = TRUE) +
  labs(x = "Aid received / GDP per capita", y = "Democracy", title = "OLS:
Democracy and aid and GDP") +
  theme_minimal() +
  scale_color_manual(values = c("dark green", "dark red")) +
  geom_vline(xintercept = (0.5 - b1[1]) / b1[2], col = "purple",
             lty = "dashed", size = 1)

## Warning: Ignoring unknown parameters: height

## `geom_smooth()` using formula 'y ~ x'
```



In the original data set the dependent variable democracy was a scale variable which ranged from 0 to 10. The 0 was the most autocratic countries and the 10 was the most democratic countries. Many countries were in between the range of 0 to 10 some being more democratic and some were more autocratic. However, this democracy scale variable is classified in to two and made a binary dummy variable here. This dummy dependent variables holds two values that 1 is democracy and 0 is non-democracy. These are plotted in the Y axis. The dark green dots represent democracy and the dark red dots represent non-democracy.

In this graph, there are two independent variables both are continuous or scale variables. One is the aid per capita and the other is Initial GDP per capita, which are plotted in the X axis. The blue line is a regression line and the straight dashed purple line is the decision boundary. The decision boundary is designed at 0.5 which is the middle value of the dependent variable democracy.

The dependent variable Y coded as 1 and 0 is regressed on X1 and X2 using linear regression and then applied a decision boundary that classify observation i as 1 if $\hat{y}i > 0.5$ and as 0 otherwise.

$yi = \beta 1 + \beta 2xi1 + \beta 3xi2 = 0.5$

From this classification regression it can be predicted that any country that holds democracy score above the threshlold value 0.5 are going to be fall in to democratic countries as non- democratic otherwise.

This multivariate regression graph shows also that there is a positive relationship between aid and democracy.

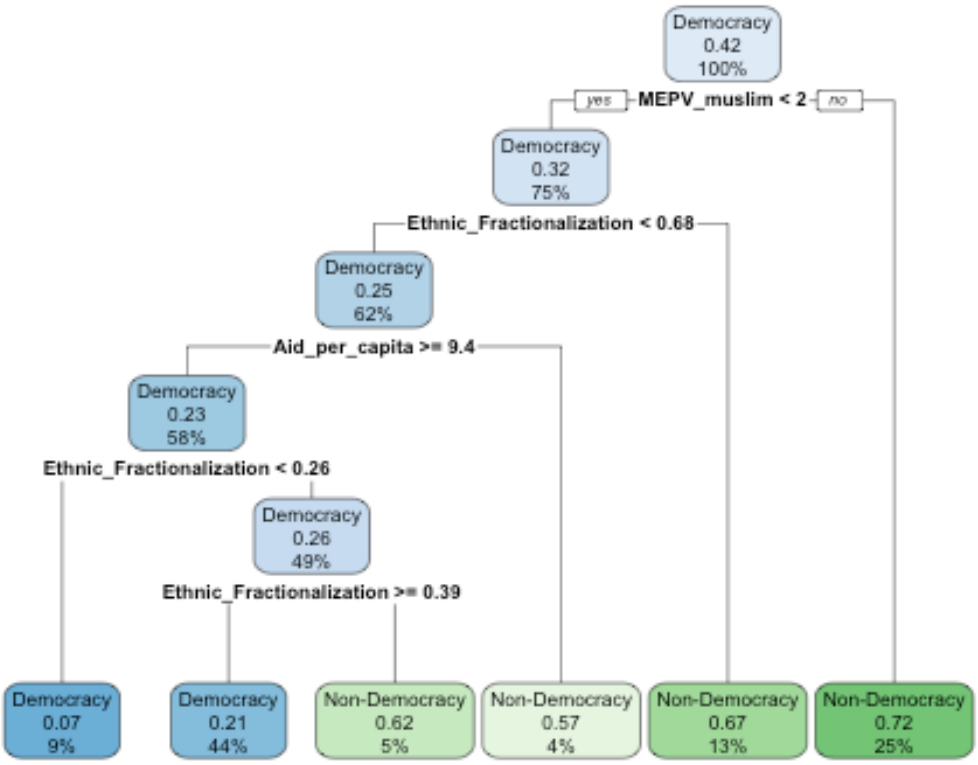
Tree based classification

DesignMetrix2 below is created for the tree based classification analysis.

```
designMatrix2 <- model.frame(democracy ~ Initial_GDP_per_capita +
Initial_population + Ethnic_Fractionalization + initial_FHdemocracy +
colonialPower + MEPV_intwar + MEPV_muslim + GDP_growth + region, data = aid)
```

Below the tree based classification is used to classify whether the countries are democratic or not by looking at the given indicators' values that are based on development and aid.

```
t <- rpart(democracy ~ Aid_per_capita +  Ethnic_Fractionalization  +
MEPV_muslim  , data = aid)
rpart.plot(t)
```



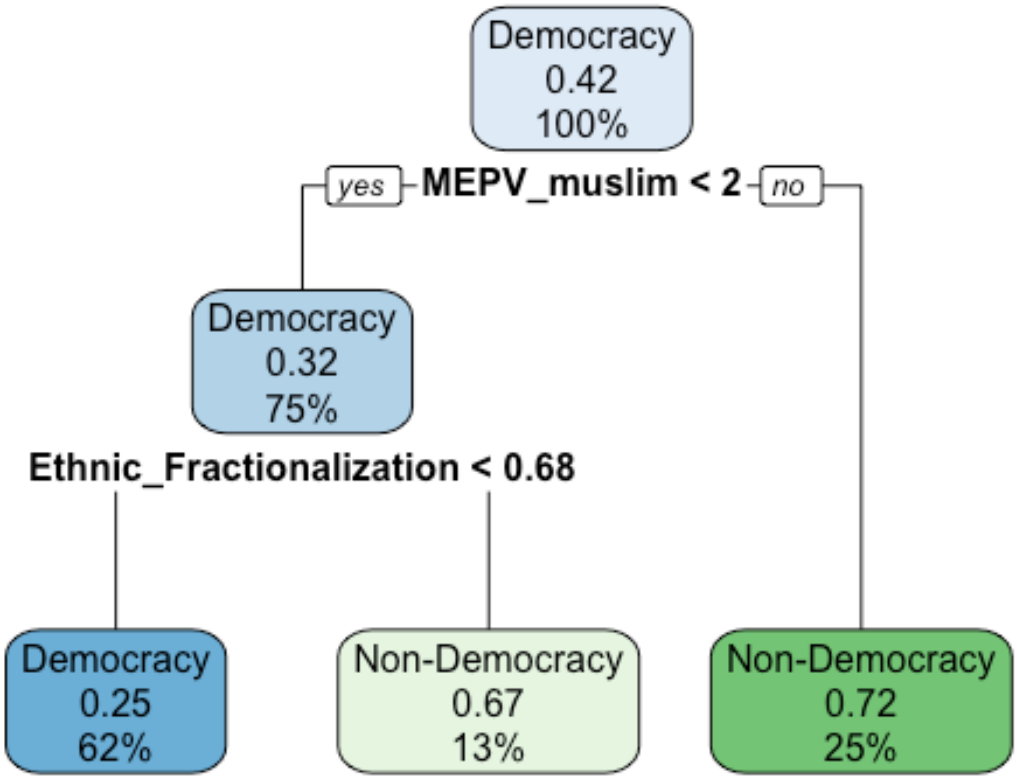
There are three indicators in this graph, which are MEPV_ Muslim, Ethnic_Fractionalization, and Aid_per_capita. The indicator MEPV Muslim data is taken from the Mepv data set, which contains political violence and conflict regions. The other two indicators are taken from the data set aid and democracy. However, all of these three indicators are being contained in designMetrix2 for the purpose of this analysis.

This graph shows that if the MEPV muslim has less than (< 2) value then it is considered to be democratic country. The scale of MEPV_Muslim value is 0 to 3 which is represents the number of political violence and conflicts in Muslim countries. Likewise, if the Ethnic Fractionalization value into less than 0.68 (< 0.68) then it also falls into the category of democracy.The Aid_per_cspita, if it is greater than or equal to 9.4 (>= 9.4) then it is considered to be democracy

Pruning tree

The above tree graph has a lot of branches, so it is vey difficult to interpret the actual outputs. Therefore to make it more interpretable the tree branches are pruned below only keeping 3 branches.

```
t3 <- prune(t, cp = .05)
rpart.plot(t3)
```



Now in this pruned tree, it is easy to explain the output. The top branch is the MEPV_Muslin and the second branch is the Ethnic_fractionalization and the third branch is the outcome of those two branches. The graph shows that if the score of Mepv Muslim is < 2 and if the score of the ethnic fractionalization is < 0.68 then they are classified as democratic countries and non-democratic otherwise.

Conclusion

In this regression and classification analysis four different data sets are used by merging and combining them together into the primary data set aid and development. Initially, the scale variable democracy is made categorical to use it as a dummy dependent variable. The relationship between aid and democracy is tested in a multiple regression, which shows a positive relationship. To see the probability prediction of democracy whether it is to be democracy or not, a multivariate graph of dummy democracy dependent variable and two independent scale variable is plotted. Father more, to see the classification of other indicators, two tree based classifications are evaluated based on the aid and development evaluation. Overall, the initial arguments that aid may promote democracy in the recipient countries remains positive.

References

Gomanee, K., Girma, S. and Morrissey, O. (2005). Aid, public spending and human welfare: evidence from quantile regressions. Journal of International Development: The Journal of the Development Studies Association, 17(3), pp.299-309.

Kersting, E. and Kilby, C. (2014) Aid and democracy redux. European Economic Review, 67, pp.125-143.

Knack, S. (2004) Does foreign aid promote democracy?. International Studies Quarterly, 48(1), pp.251-266.

Lipset, S.M.(1959) Some social requisites of democracy: Economic development and political legitimacy. The American political science review, 53(1), pp.69-105.

Data sources

Monty G. Marshall (2017) MAJOR EPISODES OF POLITICAL VIOLENCE (MEPV) AND CONFLICT REGIONS, 1946-2016 : Center for Systemic Peace <www.systemicpeace.org>

Monty G. Marshall and Gabrielle Elzinga-Marshall (2012) STATE FRAGILITY INDEX AND MATRIX, Center for Systemic Peace

Monty G. Marshall, Ted R. Gurr and Keith Jaggers (2017) POLITYTM IV PROJECT Political Regime Characteristics and Transitions, 1800-2016; Polity IV dataset version 2016 Polity is a registered trademark 3,370,976