# Socioeconomic status and Democratization <South Africa, Botswana, and Kenya compared

## Abstract

## Introduction

The dictionary meaning of "Democracy" defined by Encyclopedia Britannica is "literally ruled by the peop
Measuring democracy also has been contested that there still have ongoing debates on this subject. The r
Following the introduction, this research paper will proceed by presenting background researches about d

## Background about Democratization and its brief history in Africa

Democratization, according to An Agenda for Democratization by Boutros Boutros-Ghali, the formal Secreta
From his On Democracy, Robert Dahl suggests three conditions that are essential to attain democratic ins
Africa has also been the wave of political transitions from various types of dictatorships to more open

## Research Question and Hypotheses

The current research aims to help our understanding of the democracy with regard to its concept structu
Sub-Saharan African countries, where continuously have been experiencing democratic progress and setbac

1. There is a significant and positive correlation between socioeconomic variables and democracy.
2. Among four socioeconomic variables - GDP, primary education enrollment, gender equality, and child mortality - the educational effect on democracy is stronger than any other variables.
3. The degree of impact of variables on democracy is consistent across selected African countries.

## Literature Review

With regard to measuring democracy, Robert J. Barro, for his paper Determinants of Democracy, used the i
Further, Barro quotes Lipset's argument based on the Lipset hypothesis, which claims that increased educ
In the second literature of Democracy and Gender Equality by Caroline Beer, she contrasts the impact of
John M. Shandra et al. approaches child morality from different theoretical perspectives. By taking poli

## Data collection and data cleaning

### Data collection

In order to achieve our research purpose, firstly we have to decide what kind of data is necessary. Our reserach hypothesises are following,

- 1. There is a significant and positive correlation between socioeconomic variables and democracy.

- 2. Among four socioeconomic variables - GDP, primary education enrollment, gender equality, and child mortality - the educational effect on democracy is stronger than any other variables.
- 3. The degree of impact of variables on democracy is consistent across selected African countries.

To test these hypothesis, we need measurement of democracy and socioeconomic variables. We used following variables.

| Variable name | Detail | Source |
|---|---|---|
| polity4 | measurement of democracy | the Center for systemic peace |
| gdppc | Gross Domestic Production Per Capita | World Bank |
| pe | Primary ed enrollment | United Nations |
| mr | Child Mortality under 5 | United Nations |
| gi | gender inequality in labor market | United Nations |

- Measurement of democracy We used polity4 as a measurement of democracy. Polity4 represents the degree of democratization. The democratization level being higher, the score also being high. We downloaded from this site. http://www.systemicpeace.org/inscrdata.html
- GDP per capita:We bring this data from World Bank database. GDP per capita represents the level of satisfaction in basic needs. The following URLs are WorldBank sites in which we downloaded the time-series data of GDP for each country. http://data.worldbank.org/indicator/NY.GDP.PCAP.CD, http://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=BW, http://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=KE
- Primary education enrollment: We used Primary education enrollment as one of representative of basic needs. The URL is United Nations site in which we downloded the data. http://data.un.org/Data.aspx?q=education&d=UNESCO&f=series%3aE_1
- Mortality under 5 years old: We used this unit as measurement of health among citizens.http://data.un.org/Data.aspx?q=mortality&d=PopDiv&f=variableID%3a77
- Gender Inequality in labor market: We used gender inequality in labor market to test whether thare is another important factor other than basic needs which is suggested in modernization theory. We calculated this unit as follows,

$$GenderInequality = \frac{EmploymentRateAmongWomen}{EmploymentRateAmongMen}$$

You can find the data from ILO official site. http://www.ilo.org/ilostat/faces/help_home/data_by_subject/subject-details/indicator-details-by-subject?subject=EMP&indicator=EMP_2EMP_SEX_AGE_NB&datasetCode=YI&collectionCode=ILOEST&_afrLoop=201714902217437#!%40%40%3Findicator%3DEMP_2EMP_SEX_AGE_NB%26subject%3DEMP%26_afrLoop%3D201714902217437%26datasetCode%3DYI%26collectionCode%3DILOEST%26_adf.ctrl-state%3D102r3mzd68_271

## Data cleaning and importing into R

We downloaded each data as excel and csv file. Then, we subtructed time series data for each variables and countries and put together into three csv files which represents each countries so that we can easily import data into R. So we have three csv files each represents SouthAfrica, Botswana, and Kenya. We will import it into R.

```
dfsa <- read.csv("SA.csv", header = TRUE, sep = ",")
dfbo <- read.csv("BTW.csv", header = TRUE, sep = ",")
dfkn <- read.csv("KNY.csv", header = TRUE, sep = ",")
dfpanel <- read.csv("panel.csv", header = TRUE, sep = ",")
```

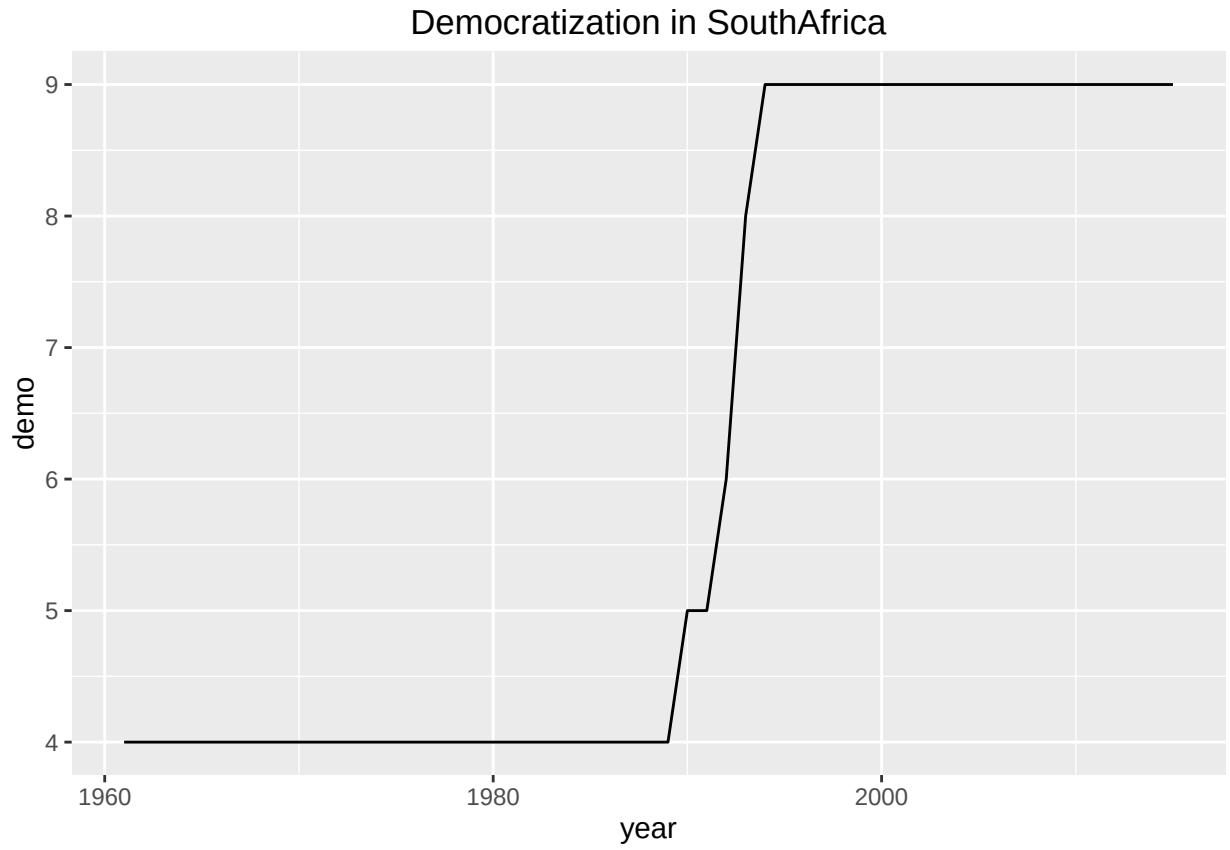
Now we suceeded to import all files.

# Descriptive Analyses

In this section, we will povide descriptive statistics of our variables.

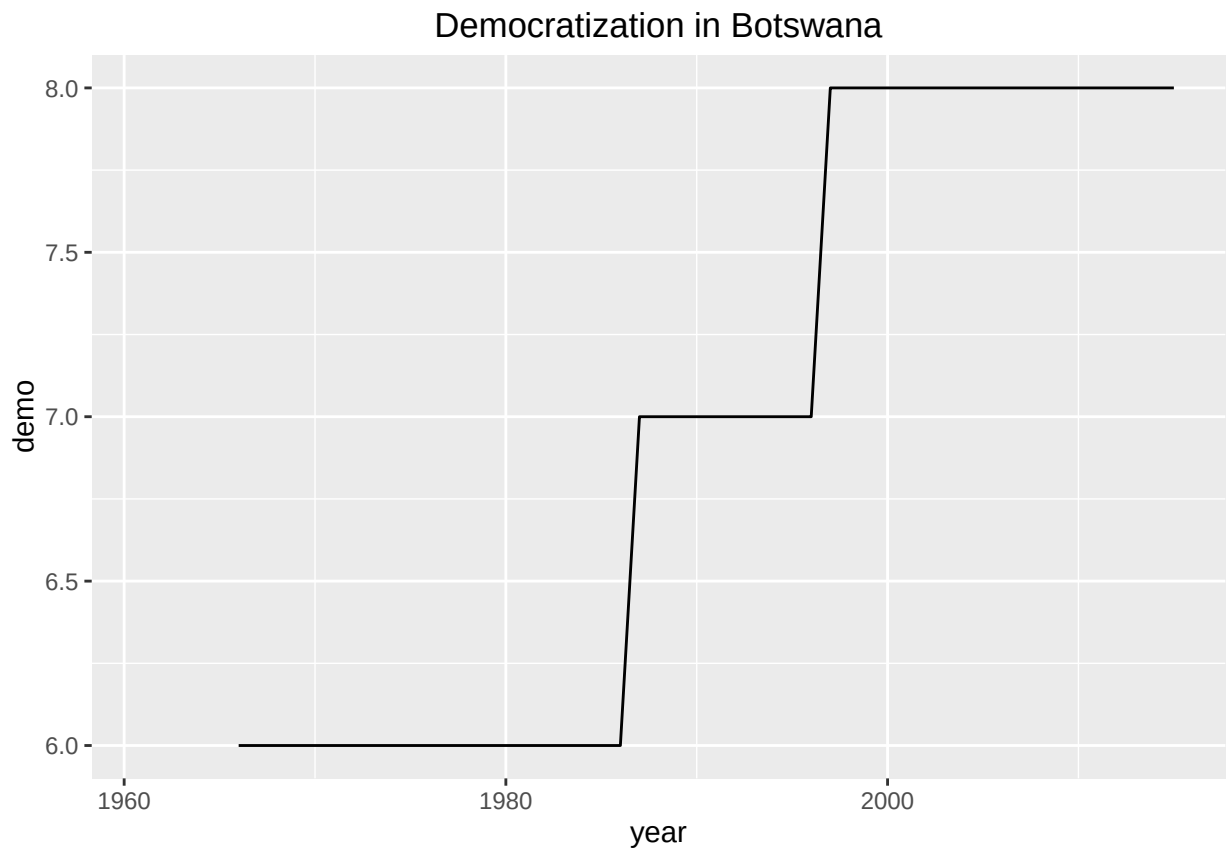
## Democratization

Following graphs are trend of democratization for each countries.

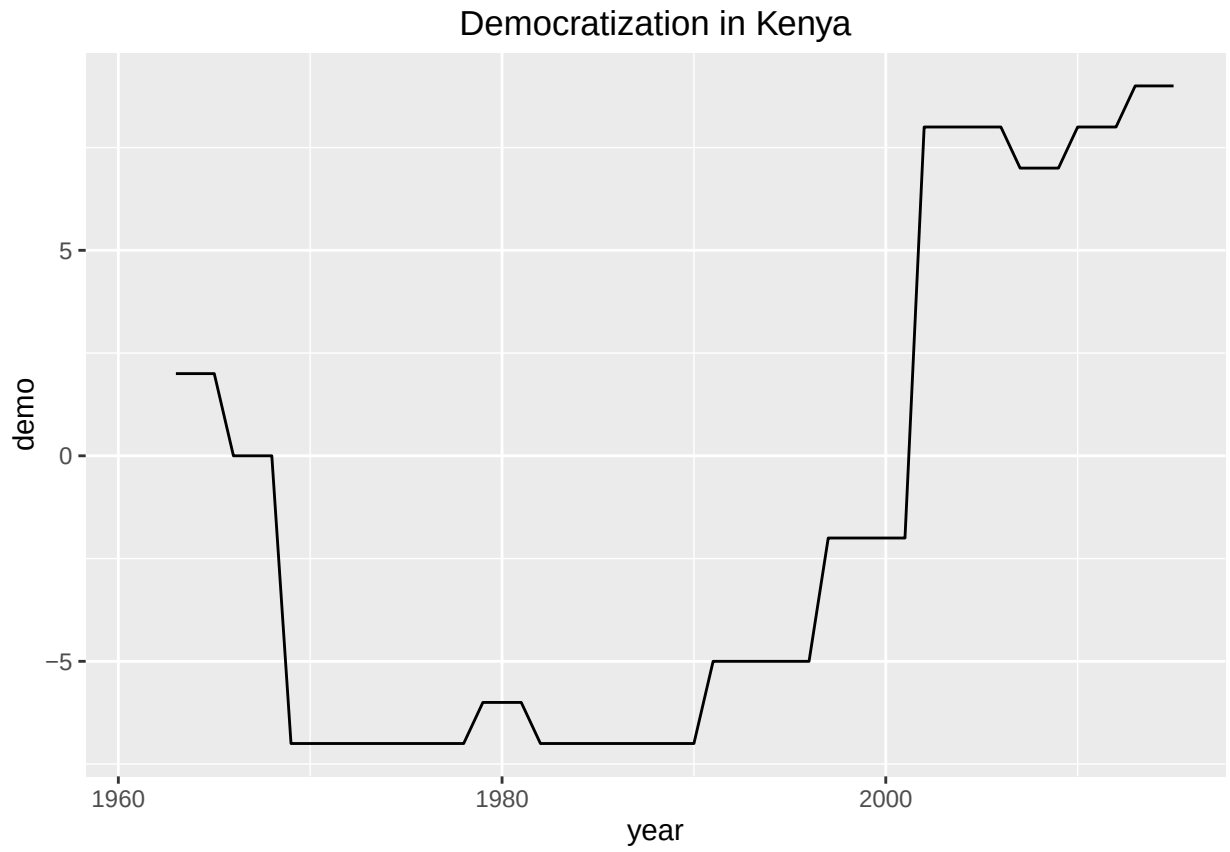



This is the trend of democratization in South Africa. As you can easily see, the level of democratization dramatically increased during 1990~1995. The contributor of this trend is the abolishment of Apartheid.

```
## Warning: Removed 5 rows containing missing values (geom_path).
```




```
## Warning: Removed 6 rows containing missing values (geom_path).
```




As graphs showing, the level of democratization has been fluctuated. Sometimes democracy advances, and sometimes it setbacks. We will investigate the data to clearify what is the driver of these fluctuation.

## Multivariate Analyses

## OLS results

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 08, 2016 - 1:34:24 PM

## Pooled OLS

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 08, 2016 - 1:34:24 PM

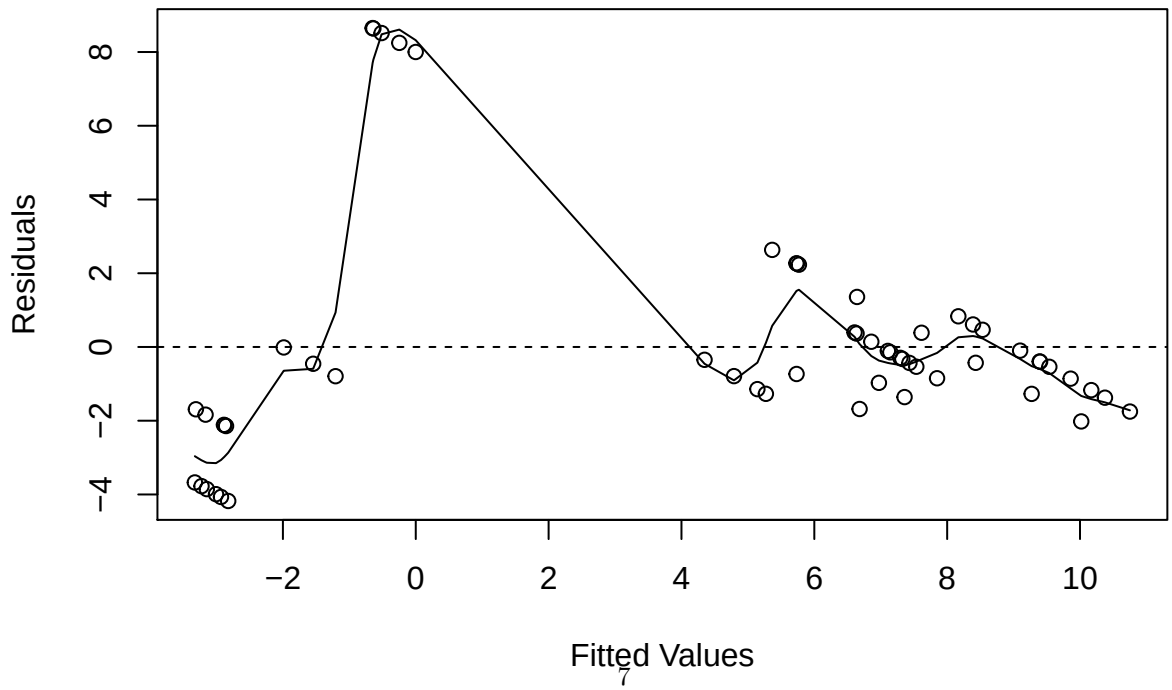
Table 2: Regression results for each country

| | *Dependent variable:* | | |
|---|---|---|---|
| | | demo | |
| | SouthAfrica | Botswana | Kenya |
| | (1) | (2) | (3) |
| log(gdppc) | $-2.78^{*}$ | $1.86^{***}$ | $-6.37$ |
| | (1.43) | (0.56) | (3.79) |
| log(pe) | 0.81 | $-3.29$ | 9.05 |
| | (2.83) | (2.31) | (8.68) |
| log(mr) | $-11.16$ | $2.84^{***}$ | $-34.05^{***}$ |
| | (7.05) | (0.84) | (9.39) |
| log(gi) | $-43.66^{***}$ | 4.61 | $-147.18^{***}$ |
| | (13.70) | (2.85) | (38.41) |
| Constant | 44.40 | 23.80 | 21.71 |
| | (73.50) | (22.73) | (157.22) |
| Observations | 17 | 21 | 19 |
| $R^2$ | 0.93 | 0.85 | 0.92 |
| Adjusted $R^2$ | 0.90 | 0.82 | 0.90 |
| Residual Std. Error | 0.72 (df = 12) | 0.28 (df = 16) | 1.94 (df = 14) |
| F Statistic | $38.99^{***}$ (df = 4; 12) | $23.04^{***}$ (df = 4; 16) | $42.66^{***}$ (df = 4; 14) |
| *Note:* | | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

Table 3: Pooled OLS

| | *Dependent variable:* |
|---|---|
| | demo |
| log(gdppc) | −0.64 |
| | (1.39) |
| log(pe) | −1.10** |
| | (0.46) |
| log(mr) | −6.89** |
| | (3.06) |
| log(gi) | −39.08*** |
| | (10.91) |
| Constant | 41.87* |
| | (21.14) |
| Observations | 57 |
| $R^2$ | 0.72 |
| Adjusted $R^2$ | 0.70 |
| Residual Std. Error | 3.16 (df = 52) |
| F Statistic | 33.01*** (df = 4; 52) |
| *Note:* | *p<0.1; **p<0.05; ***p<0.01 |

## residual vs fitted value



## Breusch-Pagan test

```
##
##  studentized Breusch-Pagan test
##
## data:  L4
## BP = 12.672, df = 4, p-value = 0.01299
```

## Fixed-Effect model

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 08, 2016 - 1:34:25 PM

Table 4: pooled OLS and fixed effects OLS

| | *Dependent variable:* | |
|---|---|---|
| | demo | |
| | PooledOLS | FixedOLS |
| | (1) | (2) |
| log(gdppc) | −0.644 | −1.496 |
| | (1.393) | (1.424) |
| log(pe) | −1.103** | 15.793*** |
| | (0.456) | (4.012) |
| log(mr) | −6.886** | −12.087*** |
| | (3.062) | (3.176) |
| log(gi) | −39.075*** | −28.940** |
| | (10.914) | (11.221) |
| Constant | 41.873* | |
| | (21.138) | |
| Observations | 57 | 57 |
| $R^2$ | 0.717 | 0.610 |
| Adjusted $R^2$ | 0.696 | 0.563 |
| F Statistic | 33.013*** (df = 4; 52) | 19.532*** (df = 4; 50) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

constants:

```
fixef(fixed)
```

```
##     Botswana        Kenya South Africa
##    -139.0599    -186.3148    -191.4030
```

## Do panel specific effects exist?

```
pFtest(fixed,pooled)
```

```
##
##  F test for individual effects
##
## data:  demo ~ log(gdppc) + log(pe) + log(mr) + log(gi)
## F = 17.075, df1 = 2, df2 = 50, p-value = 2.228e-06
## alternative hypothesis: significant effects
```

## Breush-Pagan test

```
##
##  Lagrange Multiplier Test - (Breusch-Pagan) for unbalanced panels
##
## data:  demo ~ log(gdppc) + log(pe) + log(mr) + log(gi)
## chisq = 0.67568, df = 1, p-value = 0.4111
## alternative hypothesis: significant effects
```

we cannot reject the null hypothesis. (residuals doesn't correlated with independent variables)

## Random-Effects OLS

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 08, 2016 - 1:34:26 PM

## Hausman test

```
Hausman Test
```

data: demo ~ log(gdppc) + log(pe) + log(mr) + log(gi) chisq = 4.6666e-19, df = 4, p-value = 1 alternative hypothesis: one model is inconsistent

## results

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Thu, Dec 08, 2016 - 1:34:27 PM

Oneway (individual) effect Within Model

Call: plm(formula = demo ~ log(gdppc) + log(pe) + log(mr) + log(gi), data = dfpanel, model = "within", index = c("country", "year"))

Unbalanced Panel: n=3, T=17-21, N=57

Residuals : Min. -3.160 1st Qu. -1.560 Median -0.584 3rd Qu. 0.730 Max. 8.200

Table 5: random effects OLS

| | *Dependent variable:* | | |
|---|---|---|---|
| | demo | | |
| | (1) | (2) | (3) |
| log(gdppc) | −0.644 | −1.496 | −1.496 |
| | (1.393) | (1.424) | (1.396) |
| log(pe) | −1.103** | 15.793*** | 15.793*** |
| | (0.456) | (4.012) | (3.934) |
| log(mr) | −6.886** | −12.087*** | −12.087*** |
| | (3.062) | (3.176) | (3.114) |
| log(gi) | −39.075*** | −28.940** | −28.940** |
| | (10.914) | (11.221) | (11.003) |
| Constant | 41.873* | | −172.259 |
| | (21.138) | | (997,305.100) |
| Observations | 57 | 57 | 57 |
| $R^2$ | 0.717 | 0.610 | 0.610 |
| Adjusted $R^2$ | 0.696 | 0.563 | 0.580 |
| F Statistic | 33.013*** (df = 4; 52) | 19.532*** (df = 4; 50) | 20.314*** (df = 4; 52) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 6: Regression results

| | *Dependent variable:* | | |
|---|---|---|---|
| | demo | | |
| | (1) | (2) | (3) |
| log(gdppc) | −0.644 | −1.496 | −1.496 |
| | (1.393) | (1.424) | (1.396) |
| log(pe) | −1.103** | 15.793*** | 15.793*** |
| | (0.456) | (4.012) | (3.934) |
| log(mr) | −6.886** | −12.087*** | −12.087*** |
| | (3.062) | (3.176) | (3.114) |
| log(gi) | −39.075*** | −28.940** | −28.940** |
| | (10.914) | (11.221) | (11.003) |
| Constant | 41.873* | | −172.259 |
| | (21.138) | | (997,305.100) |
| Observations | 57 | 57 | 57 |
| $R^2$ | 0.717 | 0.610 | 0.610 |
| Adjusted $R^2$ | 0.696 | 0.563 | 0.580 |
| F Statistic | 33.013*** (df = 4; 52) | 19.532*** (df = 4; 50) | 20.314*** (df = 4; 52) |
| *Note:* | | | *p<0.1; **p<0.05; ***p<0.01 |

Coefficients : Estimate log(gdppc) -1.4964 log(pe) 15.7935 log(mr) -12.0870 log(gi) -28.9397 Std. Error log(gdppc) 1.4239 log(pe) 4.0120 log(mr) 3.1755 log(gi) 11.2214 t-value log(gdppc) -1.0509 log(pe) 3.9366 log(mr) -3.8063 log(gi) -2.5790 Pr(>|t|) log(gdppc) 0.2983644 log(pe) 0.0002560 log(mr) 0.0003859 log(gi) 0.0128989

log(gdppc)
log(pe) ***log(mr)*** log(gi) *
— Signif. codes:
0 '' ***0.001*** '' *0.01* '' 0.05 '.' 0.1 '' 1

Total Sum of Squares: 788.23 Residual Sum of Squares: 307.59 R-Squared: 0.60977 Adj. R-Squared: 0.56294 F-statistic: 19.5324 on 4 and 50 DF, p-value: 9.8562e-10

$$democratization = -1.496log(gdppc) + 15.793log(pe) - 12.087log(mr) - 28.940log(gi) + \alpha_i$$

where a_i represents panel specific effects

# Conclusion

# Acknowledgment

We the authors, Takuma Andoh and Bomi Kim, would like to appreciate Professor Christopher Gandrud, discussants Pol De Santalo and Yumi Komai, and peers in Introduction to Collaborative Social Science Data Analysis for Fall 2016 at the Hertie School of Governance for their time, supports, advices, and guidance.