

# 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 people". Measuring democracy also has been contested that there still have ongoing debates on this subject. The research paper will proceed by presenting background researches about democracy in South Africa, Botswana, and Kenya.

## Background about Democratization and its brief history in Africa

Democratization, according to An Agenda for Democratization by Boutros Boutros-Ghali, the former Secretary-General of the United Nations, is a process of transition from authoritarian rule to democratic governance. From his On Democracy, Robert Dahl suggests three conditions that are essential to attain democratic institutions. Africa has also been the wave of political transitions from various types of dictatorships to more open and democratic systems.

## Research Question and Hypotheses

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

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 Polity IV index. Further, Barro quotes Lipset's argument based on the Lipset hypothesis, which claims that increased education leads to higher levels of democracy. In the second literature of Democracy and Gender Equality by Caroline Beer, she contrasts the impact of gender equality on democracy. John M. Shandra et al. approaches child mortality from different theoretical perspectives. By taking political and economic factors into account, they find that child mortality is negatively correlated with democracy.

# Data, Variables, and Methods

## Descriptive Analyses

## Multivariate Analyses

## Conclusion

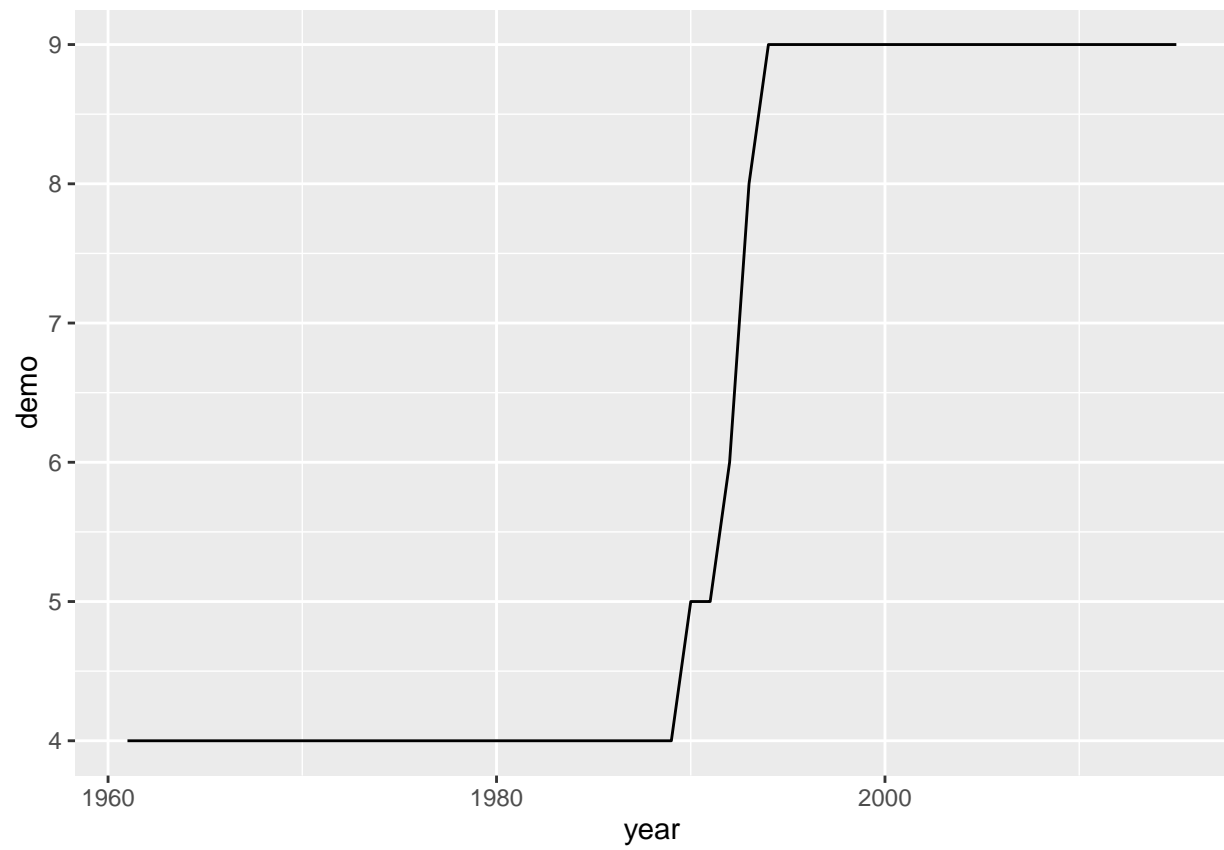
## Acknowledgment

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.

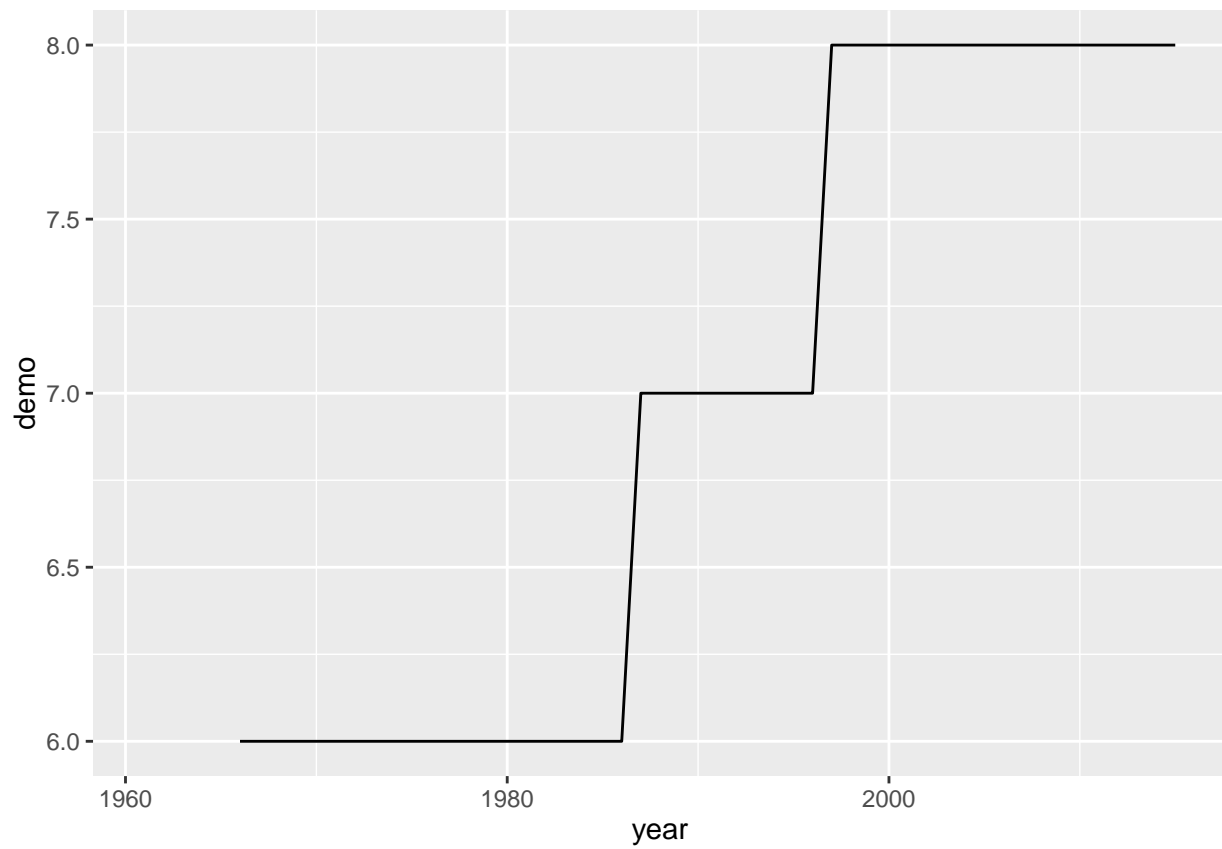
## Data Availability

- 1) Measurement of Democratization 1-1) Polity 4 ?? ??<http://www.systemicpeace.org/inscr/p4manualv2015.pdf> 1-2)
- 2) Gross National Income Level
- 3) primary enrollment
- 4) Income Inequality
- 5) Gender Inequality in labor force
- 6) mortality rate under 5

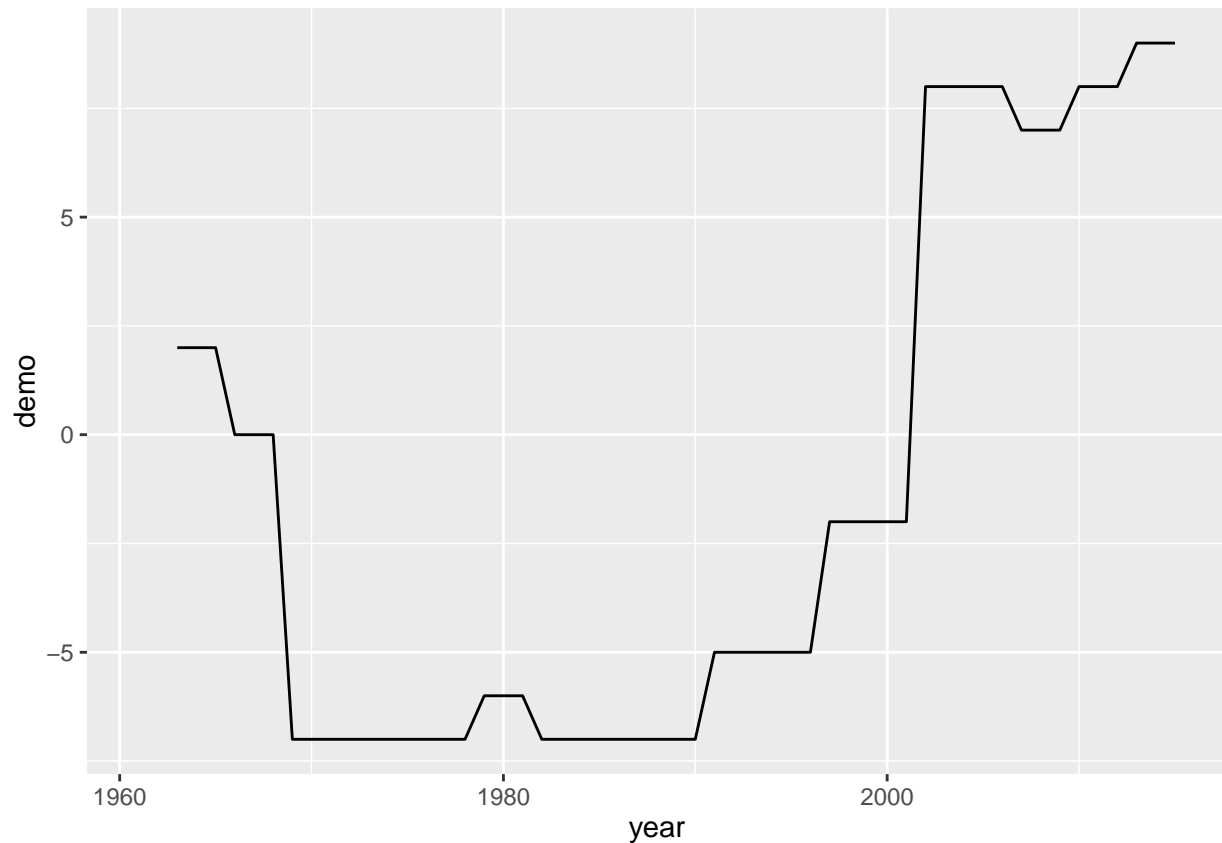
## Democratization



## Warning: Removed 5 rows containing missing values (geom\_path).



## Warning: Removed 6 rows containing missing values (geom\_path).



## Explanatory variables

variable name	detail	source
gdppc	Gross National Production Per Capita	World Bank
pe	Primary enrollment in education	United Nations
mr	Infant Mortaliry Rate	United Nations
gi	gender inequality in labor market	United Nations

## OLS results

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu  
 % Date and time: Wed, Dec 07, 2016 - 14:06:22

## Pooled OLS

% Table created by stargazer v.5.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu  
 % Date and time: Wed, Dec 07, 2016 - 14:06:22

Table 2: Regression results for each country

	<i>Dependent variable:</i>		
	SouthAfrica	demo Botswana	Kenya
	(1)	(2)	(3)
log(gdppc)	−2.78* (1.43)	1.86*** (0.56)	−6.37 (3.79)
log(pe)	0.81 (2.83)	−3.29 (2.31)	9.05 (8.68)
log(mr)	−11.16 (7.05)	2.84*** (0.84)	−34.05*** (9.39)
log(gi)	−43.66*** (13.70)	4.61 (2.85)	−147.18*** (38.41)
Constant	44.40 (73.50)	23.80 (22.73)	21.71 (157.22)
Observations	17	21	19
R <sup>2</sup>	0.93	0.85	0.92
Adjusted R <sup>2</sup>	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:*

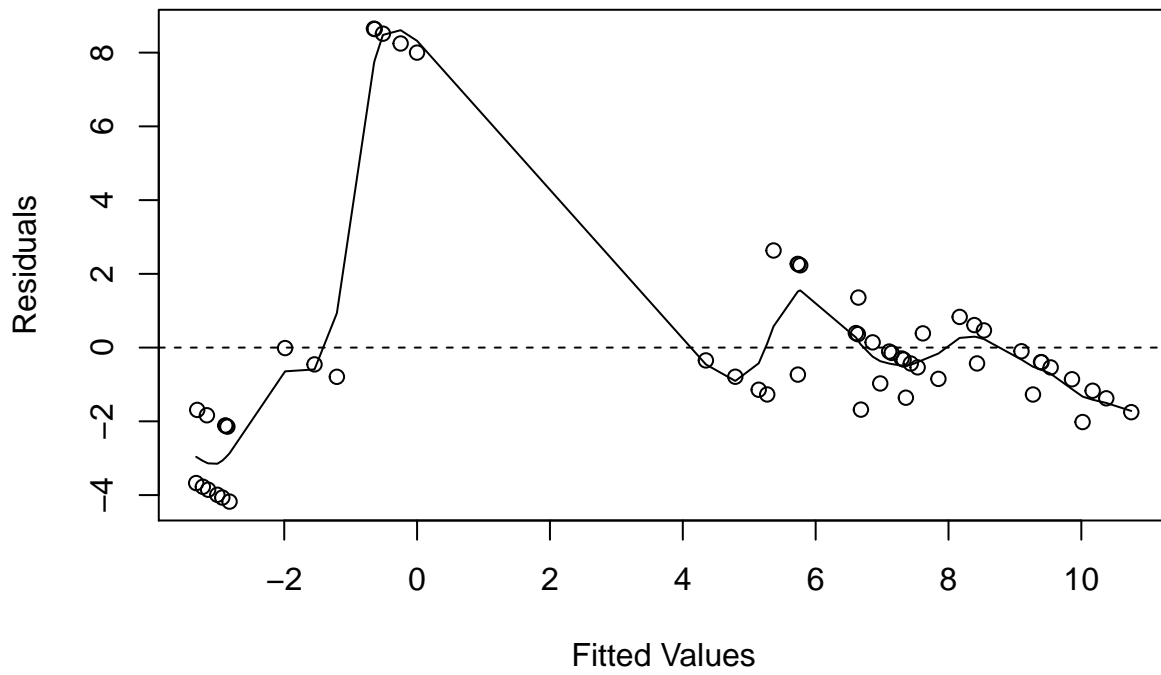
\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 3: Pooled OLS

	<i>Dependent variable:</i>
	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 <sup>2</sup>	0.72
Adjusted R <sup>2</sup>	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: Wed, Dec 07, 2016 - 14:06:23

Table 4: pooled OLS and fixed effects OLS

	<i>Dependent variable:</i>	
	demo	
	PooledOLS	FixedOLS
	(1)	(2)
log(gdppc)	−0.644 (1.393)	−1.496 (1.424)
log(pe)	−1.103** (0.456)	15.793*** (4.012)
log(mr)	−6.886** (3.062)	−12.087*** (3.176)
log(gi)	−39.075*** (10.914)	−28.940** (11.221)
Constant	41.873* (21.138)	
Observations	57	57
R <sup>2</sup>	0.717	0.610
Adjusted R <sup>2</sup>	0.696	0.563
F Statistic	33.013*** (df = 4; 52)	19.532*** (df = 4; 50)
<i>Note:</i>		
*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: Wed, Dec 07, 2016 - 14:06:23
```

## 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: Wed, Dec 07, 2016 - 14:06:24
```

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

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

Table 5: random effects OLS

	<i>Dependent variable:</i>		
	demo		
	(1)	(2)	(3)
log(gdppc)	−0.644 (1.393)	−1.496 (1.424)	−1.496 (1.396)
log(pe)	−1.103** (0.456)	15.793*** (4.012)	15.793*** (3.934)
log(mr)	−6.886** (3.062)	−12.087*** (3.176)	−12.087*** (3.114)
log(gi)	−39.075*** (10.914)	−28.940** (11.221)	−28.940** (11.003)
Constant	41.873* (21.138)		−172.259 (997,305.100)
Observations	57	57	57
R <sup>2</sup>	0.717	0.610	0.610
Adjusted R <sup>2</sup>	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&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 6: Regression results

	<i>Dependent variable:</i>		
	demo		
	(1)	(2)	(3)
log(gdppc)	−0.644 (1.393)	−1.496 (1.424)	−1.496 (1.396)
log(pe)	−1.103** (0.456)	15.793*** (4.012)	15.793*** (3.934)
log(mr)	−6.886** (3.062)	−12.087*** (3.176)	−12.087*** (3.114)
log(gi)	−39.075*** (10.914)	−28.940** (11.221)	−28.940** (11.003)
Constant	41.873* (21.138)		−172.259 (997,305.100)
Observations	57	57	57
R <sup>2</sup>	0.717	0.610	0.610
Adjusted R <sup>2</sup>	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&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

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

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

where  $\alpha_i$  represents panel specific effects

## Conclusion