

Pre-Analysis Plan

Drivers of Distributive Preferences: Material Self-Interest vs. Expressive Benefit in Africa

Ethan Christensen & Jennifer Hamilton

May 11, 2018

Introduction

Rewrite

Description of data

In this study we plan to utilize data from the World Values Survey and Afrobarometer surveys. The World Values Survey (WVS) is the largest and longest-running non-commercial cross-national survey of beliefs and values. The most recently completed wave surveyed participants from almost 100 countries in 2011-2014. Each country-level survey has a minimum of 1,200 respondents (ages 18-85), conducted through face-to-face interviews. Depending on the country, sampling occurs through either probability or a combination of probability and stratified random sampling as noted in the documentation. From these countries, we identify the five that are in sub-Saharan Africa: Ghana, Nigeria, Rwanda, South Africa, and Zimbabwe.

Afrobarometer is a long running survey designed specifically to understand political attitudes and behaviors in sub-Saharan Africa. Respondents answer questions on a variety of topics including public services, governance, identity, and political participation. There have been six completed waves, with the most recent in 2014/2015 also covering the most countries (36). Most country surveys have either 1,200 or 2,400 respondents. Each survey uses a clustered, stratified, multi-stage, area probability sample. Occasionally a survey will oversample a politically relevant sub-population in order to ensure a large-enough subsample.

From these two sources we identify the four policy preference choices available to us as listed above. In order for a policy preference question to be included in our analysis, the relevant survey also had to contain appropriate items to identify the self-interest and symbolic attitudes of respondents, hence why so few policy preference items were included.

For each topic we use different samples of countries based on the data available. The relevant questions on immigration and privatization are contained in the WVS and so is limited to those five countries. We limit our analysis of education preferences to countries where there is not compelling evidence of ethnic preference in education outcomes. If ethnic preferences exist, then material self-interest is more difficult to assess, as responses may rationally vary by ethnic group. We select countries for inclusion using Franck and Rainer's results on systematic bias in primary education attainment and literacy (2012). We limit our analysis to only countries in which no more than one of their four indicators attained statistical significance. This leaves us with five countries: Benin, Malawi, Mali, Senegal, and Uganda.

Hypotheses

H_1 : Expressive benefit indicators will be jointly significant in each model.

H_2 : Material self-interest indicators will be jointly significant in each model.

Variable construction

Immigration

Source: World Values Survey Round 6

Countries included: Ghana, Nigeria, Rwanda, South Africa, and Zimbabwe

Policy Preference

(V46) When jobs are scarce, employers should give priority to people of this country over immigrants.

- *Original coding:* 1 = agree, 2 = neither, 3 = disagree, -5 = missing, -4 = not asked, -3 = not applicable, -2 = no answer, -1 = don't know
- *Recode:* 1 = agree, 2 = neither, 3 = disagree, else NA

Expressive Benefit

1. (V107) How much you trust: People of another nationality.

- *Original coding:* 1 = trust completely, 2 = trust somewhat, 3 = do not trust very much, 4 = do not trust at all, -5 = missing, -3 = not applicable, -2 = no answer, -1 = don't know.
- *Recode:* 1 = trust completely, 2 = trust somewhat, 3 = do not trust very much, 4 = do not trust at all, else NA. *Variable type:* discrete.
- *Expected sign:* negative.

2. (V16) Important child qualities: tolerance and respect for other people.

- *Original coding:* 1 = mentioned, 2 = not mentioned, -4 = not asked, -3 = not applicable, -2 = no answer, -1 = don't know.
- *Recode:* 1 = mentioned, 0 = not mentioned, else NA. *Variable type:* binary.
- *Expected sign:* negative.

Self-Interest

1. *Personal Connection to Immigrant:* We will create a single indicator for personal connection to immigrants using the following items. If any of these items is answered as yes (1 = immigrant), personal connection to immigrant will be coded as one, else zero:

- (V243) Mother is an immigrant.
- (V244) Father is an immigrant.
- (V245) Respondent is an immigrant.

We chose to fold these items into a single index in order to overcome potential multicollinearity. *Variable type:* binary.

Expected sign: negative.

2. *Employment status of primary household breadwinner* This indicator requires use of several WVS items:

- (V229) Employment status *Original coding:* 1 = full time, 2 = part time, 3 = self-employed, 4 = retired, 5 = housewife, 6 = students, 7 = unemployed, 8 = other, -4 = not asked, -3 = not applicable, -2 = no answer, -1 = don't know.
- (V235) Are you the chief wage earner in your house? *Original coding:* 1 = yes, 2 = no, -4 = not asked, -3 = not applicable, -2 = no answer, -1 = don't know.

- (V236) Is the chief wage earner employed now or not. *Original coding:* 1 = yes, 2 = no, -4 = not asked, -3 = not applicable (Respondent is the chief wage earner), -2 = no answer, -1 = don't know.

We will code 1 if either the primary wage earner of the respondent's household (either respondent or otherwise) is employed (full time, part time, or self-employed), else 0. *Variable type:* binary. *Expected sign:* positive.

3. *Supplementary analysis:* We will include the following items in supplementary analysis. However, we exclude them from our preferred model because they are subjective measures of personal well-being, and therefore subject to endogeneity vis-a-vis value expression.

(a) (V59) Satisfaction with financial situation of household.

- *Original Coding:* scale from 1 = completely dissatisfied to 10 = completely satisfied, -4 = not asked in survey, -3 = not applicable, -2 = no answer or refused to answer, -1 = don't know.
- *Recode:* scale from 1 to 10 as above, else NA. *Variable type:* discrete.
- *Expected sign:* positive.

(b) (V181) Worries: Losing my job or not finding a job.

- *Original Coding:* 1 = very much, 2 = a great deal, 3 = not much, 4 = not much at all, -4 = not asked in survey, -3 = Inapplicable (neither has, nor seeks a job), -2 = no answer, -1 = don't know.
- *Recode:* 1 = very much, 2 = a great deal, 3 = not much, 4 = not much at all, else NA. *Variable type:* discrete.
- *Expected sign:* positive.

Controls

1. (V240) Sex, as coded by interviewer.

- *Original coding:* 1= male, 2= female, -5 = missing/unknown, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
- *Recode:* 1 = male, 0 = female, else NA. *Variable type:* binary.

2. (V241) Year of birth

- *Original coding:* decade as indicated, -5 = missing/unknown, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
- *Recode:* 1 = 1900-1909, 2 = 1910-1919, 3 = 1920-1929, 4 = 1930-1939, 5 = 1940-1949, 6 = 1950-1959, 7 = 1960-1969, 8 = 1970-1979, 9 = 1980-1989, 10 = 1990-1999, 11 = 2000-2010, else NA. *Variable type:* categorical.

3. (V248) What is the highest educational level that you have attained?

- *Original coding:* 1= no formal education; 2= incomplete primary school; 3= complete primary school; 4= incomplete secondary school: technical/vocational type; 5= complete secondary school: technical/vocational type; 6= incomplete secondary school: university-preparatory type; 7= complete secondary school: university-preparatory type; 8= some university-level education, without degree; 9= university-level education, with degree; -5 = refused to answer; -4 = not asked; -3 = not applicable; -2 = no answer; -1 = don't know.
- *Recode:* 1= no formal education; 2= incomplete primary school; 3= complete primary school; 4= incomplete secondary school: technical/vocational type; 5= complete secondary school: technical/vocational type; 6= incomplete secondary school: university-preparatory type; 7= complete secondary school: university-preparatory type; 8= some university-level education, without degree; 9= university-level education, with degree; else NA. *Variable type:* categorical.

4. (V254) Ethnic group, as coded by interviewer. As indicated. *Variable type*: categorical.
5. (V24) Most people can be trusted.
 - *Original Coding*: 1 = most people can be trusted, 2 = Need to be very careful, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
 - *Recode*: 1 = most people can be trusted, 0 = need to be very careful, else NA. *Variable type*: binary.
 - *Note*: This item is included in order to provide baseline trust in others for the item asking about trust in foreign nationals.

Privatization

Source: World Values Survey Round 6

Countries included: Ghana, Nigeria, Rwanda, South Africa, and Zimbabwe

Policy Preference

(V97) Private vs. state ownership of business

- *Original coding*: scale from 1 = Private ownership of business and industry should be increased to 10 = Government ownership of business and industry should be increased, -5 = missing, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
- *Recode*: scale from 1 = Private ownership of business and industry should be increased to 10 = Government ownership of business and industry should be increased, else NA.

Expressive Benefit

1. (V96) Income inequality ideology.
 - *Original coding*: scale from 1 = Incomes should be made more equal to 10 = We need larger income differences as incentives for individual effort, -5 = missing, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
 - *Recode*: scale from 1 = Incomes should be made more equal to 10 = We need larger income differences as incentives for individual effort, else NA. *Variable type*: discrete.
 - *Expected sign*: negative.
2. (V98) Government responsibility.
 - *Original Coding*: scale from 1 = Government should take more responsibility to ensure that everyone is provided for to 10 = People should take more responsibility to provide for themselves, -5 = missing, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
 - *Recode*: scale from 1 = Government should take more responsibility to ensure that everyone is provided for to 10 = People should take more responsibility to provide for themselves, else NA. *Variable type*: discrete.
 - *Expected sign*: negative.

Self-interest

1. (V230) Sector of employment.
 - *Original coding*: 1 = government or public institution, 2 = private business or industry, 3 = private non-profit organization, 4 = autonomous/informal sector/other, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.

- *Recode*: 1 = government or public institution; 0 = private business or industry, private non-profit organization, autonomous/informal sector/other; else NA. *Variable type*: binary.
- *Expected sign*: positive.
- *Note*: this indicator is included on the basis of previous research on preferences over privatization [battaglio_self-interest_2009].

Controls

1. (V240) Sex, as coded by interviewer.
 - *Original coding*: 1= male, 2= female, -5 = missing/unknown, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
 - *Recode*: 1 = male, 0 = female, else NA. *Variable type*: binary.
2. (V241) Year of birth
 - *Original coding*: decades as indicated, -5 = missing/unknown, -4 = not asked in survey, -3 = not applicable, -2 = no answer, -1 = don't know.
 - *Recode*: 1 = 1900-1909, 2 = 1910-1919, 3 = 1920-1929, 4 = 1930-1939, 5 = 1940-1949, 6 = 1950-1959, 7 = 1960-1969, 8 = 1970-1979, 9 = 1980-1989, 10 = 1990-1999, 11 = 2000-2010, else NA. *Variable type*: categorical.
3. (V248) What is the highest educational level that you have attained?
 - *Original coding*: 1= no formal education; 2= incomplete primary school; 3= complete primary school; 4= incomplete secondary school: technical/vocational type; 5= complete secondary school: technical/vocational type; 6= incomplete secondary school: university-preparatory type; 7= complete secondary school: university-preparatory type; 8= some university-level education, without degree; 9= university-level education, with degree; -5 = refused to answer; -4 = not asked; -3 = not applicable; -2 = no answer; -1 = don't know.
 - *Recode*: 1= no formal education; 2= incomplete primary school; 3= complete primary school; 4= incomplete secondary school: technical/vocational type; 5= complete secondary school: technical/vocational type; 6= incomplete secondary school: university-preparatory type; 7= complete secondary school: university-preparatory type; 8= some university-level education, without degree; 9= university-level education, with degree; else NA. *Variable type*: categorical.
4. (V254) Ethnic group, as coded by interviewer. As indicated. *Variable type*: categorical.

School fees

Source: Afrobarometer Round 3

Countries included: Benin, Malawi, Mali, Senegal, Uganda

Policy Preference

(Q10) Which of the following statements is closest to your view?. Choose Statement A or Statement B. A: It is better to have free schooling for our children, even if the quality of education is low. B: It is better to raise educational standards, even if we have to pay school fees.

- *Original Coding*: 1=Agree Very Strongly with A, 2=Agree with A, 3=Agree with B, 4=Agree Very Strongly with B, 5=Agree with Neither, 9=Don't Know, 98=Refused to Answer, -1=Missing Data
- *Recode*: 1=Agree Very Strongly with A, 2=Agree with A, 3=Agree with Neither, 4=Agree with B, 5=Agree Very Strongly with B, else NA.

Expressive Benefit

1. (Q19) Let's talk for a moment about the kind of society we would like to have in this country. Which of the following statements is closest to your view?. Choose Statement A or Statement B. A: People should look after themselves and be responsible for their own success in life. B: The government should bear the main responsibility for the well-being of people.
 - *Original Coding:* 1=Agree Very Strongly with A, 2=Agree with A, 3=Agree with B, 4=Agree Very Strongly with B, 5=Agree with Neither, 9=Don't Know, 98=Refused to Answer, -1=Missing Data.
 - *Recode:* 1=Agree Very Strongly with A, 2=Agree with A, 3=Agree with Neither, 4=Agree with B, 5=Agree Very Strongly with B, else NA. *Variable type:* discrete.
 - *Expected sign:* negative.

Self-interest

Afrobarometer asks two questions in separate sections regarding the affordability of school expenses and fees. We include the first in our preferred specification. However, we will run models with the second item as well for a robustness check.

1. *Public School Contact* Our self-interest items rely on having direct experience with the public schooling system. As such, we code a term for whether the respondent indicates having contact with the public school system. On each of the items listed below (Q73A-F), respondents have the option to respond "no experience with public schools in the past twelve months" (coded as 7). Our new variable is coded 0 if the respondent indicates "no experience" for ANY of the questions Q73A-F, else 1. We will interact this variable with each of the self interest variables below, effectively turning it "on" or "off" according to contact.

We do observe that the percentage of respondents choosing "no experience" sometimes vary from question to question. This is concerning, as respondents with any experience with the public school system should answer "don't know" or "never" rather than "no experience" for these questions if they have not encountered the problem listed. Because the responses of those who answer "no experience" for some questions but not others is ambiguous, we include all of these in the no contact category. *Variable type:* binary.

2. (Q73A) Have you encountered any of these problems with your local public schools during the past 12 months?. Services are too expensive / Unable to pay.
 - *Original Coding:* 0=Never, 1=Once or twice, 2=A few times, 3=Often, 7=No experience with public schools in the past twelve months, 9=Don't Know, 98=Refused to Answer, -1=Missing Data.
 - *Recode:* 0=Never, 1=Once or twice, 2=A few times, 3=Often, 0=No experience with public schools in the past twelve months, else NA. *Variable type:* discrete.
 - *Expected sign:* negative.
3. *School quality index* We will construct an additive index of school quality concerns, based on the following items:
 - (a) (Q73B) Have you encountered any of these problems with your local public schools during the past 12 months?. Lack of textbooks or other supplies.
 - (b) (Q73C) Have you encountered any of these problems with your local public schools during the past 12 months?. Poor teaching.
 - (c) (Q73D) Have you encountered any of these problems with your local public schools during the past 12 months? Absent teachers.
 - (d) (Q73E) Have you encountered any of these problems with your local public schools during the past 12 months?. Overcrowded classrooms.

- (e) (Q73F) Have you encountered any of these problems with your local public schools during the past 12 months? Poor conditions of facilities.

Each of these questions has identical response options:

- *Original coding:* 0=Never, 1=Once or twice, 2=A few times, 3=Often, 7=No experience with public schools in the past twelve months, 9=Don't Know, 98=Refused to Answer, -1=Missing Data.
- *Recode:* 0=Never, 1=Once or twice, 2=A few times, 3=Often, 0=No experience with public schools in the past twelve months, else NA.
Response for each Q73B-F will be summed. *Variable type:* discrete.

Expected sign: positive.

Controls

1. (Q90) What is the highest level of education you have completed?
 - *Original coding:* 0=No formal schooling, 1=Informal schooling (including Koranic schooling), 2=Some primary schooling, 3=Primary school completed, 4=Some secondary school/ High school, 5=Secondary school completed/High school, 6=Post-secondary qualifications, other than university e.g. a diploma or degree from a technical/polytechnic/college, 7=Some university, 8=University completed, 9=Post-graduate, 98=Refused to Answer, 99=Don't Know, -1=Missing Data.
 - *Recode:* 0=No formal schooling, 1=Informal schooling (including Koranic schooling), 2=Some primary schooling, 3=Primary school completed, 4=Some secondary school/ High school, 5=Secondary school completed/High school, 6=Post-secondary qualifications, other than university e.g. a diploma or degree from a technical/polytechnic/college, 7=Some university, 8=University completed, 9=Post-graduate, else NA. *Variable type:* categorical.
2. (Q101) Respondent's gender, as assessed by interviewer.
 - *Original coding:* 1=Male, 2=Female
 - *Recode:* 1 = male, 0 = female. *Variable type:* binary.
3. (Q116B) Were the following services present in the primary sampling unit/enumeration area: School?
 - *Original coding:* 0=No, 1=Yes, 9=Can't determine, -1=Missing Data.
 - *Recode Scaling:* 0=No, 1=Yes, else NA. *Variable type:* binary.
4. (Q1) How old are you?
 - *Original coding:* 18-110 as indicated, 998=Refused to Answer, 999=Don't Know, -1=Missing Data.
 - *Recode:* 18-110 as indicated, else NA. *Variable type:* discrete.
5. (Q79) What is your tribe? You know, your ethnic or cultural group.
 - *Original coding:* : see AFB for full list, 990=National identity only, 995=Other, 997=Not Asked, 998=Refused, 999=Don't know, -1=Missing Data.
 - *Recode:* As above, except NA for original codes 997, 998, 999, -1. *Variable type:* categorical.
6. (Q81) How often are [respondent's identity group] treated unfairly by the government?
 - *Original Coding:* 0=Never, 1=Sometimes, 2=Often, 3=Always, 7=Not Applicable, 9=Don't Know, 98=Refused to Answer, -1=Missing Data
 - *Recode:* 0=Never, 1=Sometimes, 2=Often, 3=Always, else NA. *Variable type:* discrete.

- *Note:* Although we focus on countries without a history of ethnic bias in education, political preferences may nonetheless be driven by perceived ethnic bias, either in education or in other domains. As such, we include this measure to mitigate omitted variable bias vis-a-vis ethnically driven policy preference.

Model

For each policy area, we will use an ordered logit model with policy preference as the dependent variable and with symbolic attitudes and self-interest indices as the independent variables. We will also include control variables as indicated above.

Immigration and Privatization Models

These model will take the form:

$$y_i = \alpha + \mathbf{X}_{i1}\beta + \mathbf{X}_{i2}\gamma + \mathbf{X}_{i3}\delta + \pi_i + \epsilon_i \quad (1)$$

where:

- \mathbf{X}_{i1} represents a matrix of expressive benefit indicators and β represents a vector of coefficients for each expressive benefit indicator.
- \mathbf{X}_{i2} represents a matrix of self-interest indicators and γ represents a vector of coefficients for each self-interest indicator.
- \mathbf{X}_{i3} represents a matrix of control variables and δ represents a vector of coefficients for each control variable.
- π_i represents country fixed effects.

Note that symbolic attitudes and self-interest are thus *not* summary indices (except in cases indicated under variable construction); rather, each indicator is entered individually into the model.

School Fees Model

This model will take the form:

$$y_i = \alpha + \mathbf{X}_{i1}\beta + \mathbf{X}_{i2} \cdot X_{i4}\gamma + \mathbf{X}_{i2}\delta + \mathbf{X}_{i4}\zeta + X_{i3}\theta + \pi_i + \epsilon_i \quad (2)$$

where:

- \mathbf{X}_{i1} represents a matrix of expressive benefit indicators and β represents a vector of coefficients for each expressive benefit indicator.
- \mathbf{X}_{i2} represents a matrix of self-interest indicators, X_{i4} represents a binary indicator for school contact, and $\mathbf{X}_{i2} \cdot X_{i4}$ represents interactions between each self-interest indicator and school contact. γ represents a vector of coefficients for the interaction terms, δ represents the vector of coefficients for self interest terms alone, and ζ represents the coefficient of the school contact indicator alone.
- \mathbf{X}_{i3} represents a matrix of control variables, and θ represents a vector of coefficients for each control variable.
- π_i represents country fixed effects.

Note that symbolic attitudes and self-interest are thus *not* summary indices (except in cases indicated under variable construction); rather, each indicator is entered individually into the model.

Robustness check

As an alternative to the interactive model used above, we will also conduct a robustness check in which we subset to only respondents with a 1 for school contact, and use model (1) estimate effects and test relevant hypotheses.

Hypothesis Testing

For each model, we will conduct two joint F-test. The first will consider whether expressive benefits indicators are jointly significant, and the second will consider whether material self-interest indicators are jointly significant.

Expressive Benefits

The F-test for joint significance of expressive benefits indicators has the following hypotheses:

1. For Model 1 (immigration, privatization):

- $H_0: \gamma = \mathbf{0}$
- $H_A: \beta \neq \mathbf{0}$

Hence, the restricted model will be $y_i = \alpha + \mathbf{X}_{i1}\beta + \mathbf{X}_{i3}\delta + \pi_i + \epsilon_i$.

2. For Model 2 (school fees):

- $H_0: \beta = \mathbf{0}$
- $H_A: \beta \neq \mathbf{0}$

Hence, the restricted model will be $y_i = \alpha + \mathbf{X}_{i1}\beta + \mathbf{X}_{i2} \cdot X_{i4}\gamma + \mathbf{X}_{i2}\delta + \mathbf{X}_{i4}\zeta + X_{i3}\theta + \pi_i + \epsilon_i$.

Self-interest

The F-test for joint significance of self-interest indicators has the following hypotheses:

1. Model 1 (immigration, privatization):

- $H_0: \beta_2 = \mathbf{0}$
- $H_A: \beta_2 \neq \mathbf{0}$

Hence, the restricted model will be $y_i = \alpha + \mathbf{X}_{i1}\beta + X_{i3}\theta + \pi_i + \epsilon_i$.

2. For Model 2 (school fees):

- $H_0: \gamma = \delta = \zeta = \mathbf{0}$
- $H_A: \text{at least one}\{\gamma, \delta, \zeta\} \neq 0$

Hence, the restricted model will be $y_i = \alpha + \mathbf{X}_{i2}\gamma + \mathbf{X}_{i3}\delta + \pi_i + \epsilon_i$.

Standard errors

For both joint F-test, we will use cluster-robust standard errors, with primary sampling units as the cluster.

Procedures for Missing Data

We will consider the problem of missing data separately for each model. In each model, if less than 5% of observations contain missing data on at least one covariate, we will simply drop missing observations. If 5% or more of observations contain missing data on at least one covariate, we will impute missing data.

Multiple hypothesis testing

We will use the Benjamini-Hochberg Procedure to address false discovery through multiple hypothesis testing. In other words, we will:

- Order the p-values for each of m hypothesis tests such that $p_1 \leq p_2 \leq \dots \leq p_m$.
- Maximize k such that $p_i \leq \frac{\alpha \cdot k}{m}$. This value will be k^* .
- Reject all hypotheses H_k for $k \leq k^*$

References

FRANCK, RAPHAËL, and ILIA RAINER. 2012. “Does the Leader’s Ethnicity Matter? Ethnic Favoritism, Education, and Health in Sub-Saharan Africa.” *The American Political Science Review* 106 (2): 294–325.