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

Based on the original findings of the Sears and coauthors' papers, we expect that both expressive benefits and and self-interest will be jointly significant in the model for each question considered.

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.
    - *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.
    - *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. *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 (Repondent 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 in any capacity (full time, part time, or self-employed), else 0. *Expected sign*:

3. *Supplementary analysis*: We will include the following items in supplementary anaylsis. However, we exclude them from our prefered 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
   - *Expected sign*:

   (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.
   - *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.
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.
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.
4. (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.

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.
    - *Expected sign*: negative.
2. (V98) Government responsibility.
    - *Original Coding*: scale from 1 = Government should take more responsiboity to ensure that everone 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 responsiboity to ensure that everone is provided for to 10 = People should take more responsibility to provide for themselves, else NA.
    - *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.
    - *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.

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.
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.
4. (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.

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.

- *Expected sign*: negative.

Self-interest

1. (Q8F) Over the past year, how often, if ever, have you or your family gone without: School expenses for your children (like fees, uniforms or books)?
   - *Original Coding*: 0=Never, 1=Just once or twice, 2=Several times, 3=Many times, 4=Always, 7=No children, 9=Don't Know, 998=Refused to Answer, -1=Missing Data
   - *Recode*:
   - *Expected sign*: negative.
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*:
   - *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*:

   We expect a sdfl;asdfl

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.

2. (Q101) Respondent's gender, as assessed by interviewer.
   - *Original coding*: 1=Male, 2=Female
   - *Recode*: 1 = male, 0 = female
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.
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
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*:

Also considering

1. Think about the condition of [respondent's identity group]. Do they have less, the same, or more influence in politics than other groups in this country? *Scaling*: (1) much more, more, same, less, much less (0). *Expected sign*: positive.
2. How often are [respondent's identity group] treated unfairly by the government? *Scaling*: (1) never, sometimes, often, always (0). *Expected sign*: positive.
3. Let us suppose that you had to choose between being a [nationality] and being a [respondent's identity group]. Which of the following statements best expresses your feelings? *Scaling*: (1) ethnic ID only, ethnic ID more, national and ethnic, national ID more, national ID only (0). *Expected sign*: negative.
4. How much do you trust each of the following types of people: [countrywomen] from other ethnic groups? *Scaling*: (1) Not at all, just a little, somewhat, a lot (0). *Expected sign*: negative.

Model

Note: We are still deciding which model would be most appropriate, as response variables are not binary but they are bounded from 0 to 1. Logit might be better? We will determine this before finalizing our PAP.

For each policy area, we will run an OLS regression with policy preference as the dependent variable and with symbolic attitudes and self-interest indices as the independent variables. We will also control for demographic variables including gender, age, education, and ethnicity. Thus, the model will take the form:

$$y_i = \beta_0 + \mathbf{X}_{i1}\beta_1 + \mathbf{X}_{i2}\beta_2 + \mathbf{X}_{i3}\beta_3 + \epsilon_i \tag{1}$$

where:

- $\mathbf{X}_{i1}$ represents a matrix of symbolic attitudes indicators and β_1 represents a vector of coefficients for each symbolic attitude indicator.
- $\mathbf{X}_{i2}$ represents a matrix of self-interest indicators and β_2 represents a vector of coefficients for each self-interest indicator.
- $\mathbf{X}_{i3}$ represents a matrix of control variables and β_3 represents a vector of coefficients for each control variable.

Note that symbolic attitudes and self-interest are thus *not* summary indices; rather, each indictor is entered individually into the model. This replicates the original methodology used in Sears and coauthors' series of symbolic politics papers (Sears, Hensler, and Speer 1979; Sears et al. 1980).

Hypothesis tests

For each model, we will conduct the following hypothesis tests:

Joint F tests

To test whether symbolic attitudes and material self-interest each contribute to the formation of policy preferences, we will use joint F-tests. These test are NOT drawn from the Sears and coauthors papers, which only consider the statistical significance of each covariate individually.

Symbolic Attitudes

Recalling that β_1 is a vector of coefficients on all symbolic attitude covariates, the F-test for joint significance of symbolic attitudes indicators has the following hypotheses:

$$\begin{aligned} H_0 &: \beta_1 = \mathbf{0} \\ H_A &: \beta_1 \neq \mathbf{0} \end{aligned}$$

Hence, the restricted model will be

$$y_i = \beta_0 + \mathbf{X}_{i2}\beta_2 + \mathbf{X}_{i3}\beta_3 + \epsilon_i \tag{2}$$

Self-interest

Recalling that β_2 is a vector of coefficients on all self-interest covariates, the F-test for joint significance of self-interest indicators has the following hypotheses:

$$\begin{aligned} H_0 &: \beta_2 = \mathbf{0} \\ H_A &: \beta_2 \neq \mathbf{0} \end{aligned}$$

Hence, the restricted model will be

$$y_i = \beta_0 + \mathbf{X}_{i1}\beta_1 + \mathbf{X}_{i3}\beta_3 + \epsilon_i \tag{3}$$

Standard errors

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

R-squared comparison test

In the original Sears and coauthors analysis compared the relative contribution of symbolic attitudes and self-interest to policy preferences by comparing the R^2 contribution made by each set of covariates. For instance, the contribution of symbolic attitudes will be:

$$R^2_{\text{symbolic}} = R^2_1 - R^2_2$$

where

- R^2_{symbolic} represents the explanatory contribution of symbolic attitudes.
- R^2_1 represents the coefficent of determination for the full model, including symbolic attitude indicators, self-interest indicators, and control variables (equation 1).
- R^2_2 represents the coefficient of determination for the restricted model which excludes symbolic attitude indicators, but includes both self-interest indicators and control variables (equation 2).

Using the same technique, the contribution of self-interest will be:

$$R^2_{\text{self-interest}} = R^2_1 - R^2_3$$

where

- $R^2_{\text{self-interest}}$ represents the explanatory contribution of self interest.
- R^2_1 represents the coefficent of determination for the full model, including symbolic attitude indicators, self-interest indicators, and control variables (equation 1).
- R^2_2 represents the coefficient of determination for the restricted model which excludes self-interest indicators, but includes both symbolic attitudes indicators and control variables (equation 2).

We will also use these estimates. Because the number of items for self-interest and symbolic attitudes varies, we will use adjusted R^2 in all analysis; for instance, we would not want to bias our results simply because the battery of symbolic attitude items is longer than than the battery of self-interest items. In addition, we will extend the original analysis by estimating the standard errors of the R^2 contributions using bootstrapping, enabling us to assess more rigorously whether contribution of either symbolic attitudes or self-interest clearly outweighs the other. Thus, we will test the following hypotheses:

$$H_0 : R^2_{\text{symbolic}} - R^2_{\text{self-interest}} = 0$$
$$H_A : R^2_{\text{symbolic}} - R^2_{\text{self-interest}} \neq 0$$

We will conduct 10,000 bootstraps to estimate standard errors. Our goal is to maintain consistency between the original sampling procedure and our bootstrap resampling method. However, we face certain practical limitations.

Disclaimer: We are not yet settled on our bootstrapping procedure and will seek further advice before finalizing our PAP.

Afrobarometer uses consistent sampling policies across countries and across survey rounds. For Afrobaromter questions, we will (1) divide the data back into country samples, (2) stratify the data by urban and rural PSUs within each country, and (3) resample PSUs within each strata. The World Values Survey does not use consistent sampling methods across countries. As such, we resample according to the lowest common denominator among country sampling procedures (1) divide the data back into country samples, and (2) resample PSUs within each country.

Multiple hypothesis testing

We will use the Benjamimi-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 $p_1 \leq p_2 \leq \ldots \leq p_m$.
- 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.

Sears, David O., Carl P. Hensler, and Leslie K. Speer. 1979. "Whites' Opposition to 'Busing': Self-Interest or Symbolic Politics?" *The American Political Science Review* 73 (2): 369–84. doi:10.2307/1954885.

Sears, David O., Richard R. Lau, Tom R. Tyler, and Harris M. Allen. 1980. "Self-Interest Vs. Symbolic Politics in Policy Attitudes and Presidential Voting." *The American Political Science Review* 74 (3): 670–84. doi:10.2307/1958149.