NYU | WAGNER

Alec Gilfillan
Advanced Quantitative Analysis II
May 10, 2018
Final Assignment

Summary: This memo uses public opinion data to examine the relationship between the number of languages spoken in a country and the self-rated quality of life in comparison to fellow citizens. An ordered logit model is used to show that survey respondents were more likely to rate their living conditions as being worse than other citizens in countries as the total number of languages in the country increased.

Data: Afrobarometer is a non-partisan research network that conducts public opinion surveys in Africa. The data for this assignment was downloaded from http://afrobarometer.org on May 7, 2018 and converted to data for Stata 14 via SPSS output. The most recent round of data was used (round 6), which covers 36 countries in 2016.

Research Question: Language is an important identifier for cultural groups. Of the 36 countries in the Afrobarometer survey, there is wide range of languages spoken. All 1,198 respondents from Egypt reported speaking Egyptian Arabic and home, while Tanzania's 2,386 respondents reported speaking 52 different languages. The research question in this memo is how this wide range of languages spoken in a country has an effect on perceptions of self and others.

Question 5 in the survey asks "In general, how do you rate your living conditions compared to those of other [ENTER NATIONALITY]?" with the possible responses being:

1. Much worse
2. Worse
3. Same
4. Better
5. Much better
6. Don't Know
7. Refused to answer

Missing data: There were 53,935 responses to question 5. Around 2.5% of the responses were "Don't know" or "Refused to answer." Because of the large sample size, responses 6 and 7 were dropped from the analysis.

Descriptive Statistics:

The responses to question 5 are tabulated in Table 1. Respondents are generally more positive than negative and the majority are neutral. The lowest category of response is "Much worse" with 9.7%.

Table 1

Question 5 Frequencies - All Countries

Q5: Your living conditions vs. others	Frequency	Percent
Much Worse	5,086	9.67
Worse	13,842	26.33
Same	17,306	32.92
Better	14,430	27.45
Much Better	1,913	3.65
Total	52,577	

Method: An ordered logit regression was chosen because it is a common analysis for survey data where there is an order to the responses but with no clear, meaningful distance between them.

Question 2 of the survey asks the respondents which language they speak at home. A continuous variable was created for each observation which contains the total number of languages spoken at home in each country.

Controls: The survey includes age (a continuous variable), country (categorical), and binary indicators for gender and for the respondent's home being in an urban location.

Model 1: An ordered logit regression with responses #1-5 as the categorical dependent variable, total number of languages spoken as the independent variable, and controls for age and country.

Model 2: An ordered logit regression with recoded responses for question 5 (1: worse 2: same 3: better) as the categorical dependent variable, total number of languages spoken as the independent variable, and controls for age and country.

Model 3: An ordered logit regression with recoded responses for question 5 (1: worse 2: same 3: better) as the categorical dependent variable, total number of languages spoken as the independent variable with an interactive term for total number of languages squared, and controls for age, country, gender, and urban location.

Table 2 compares all three models while showing statistical significance. The variable of interest, total number of languages spoken, is significant in all three. In model 3, the dummy variable for urban location has a very large significant effect.

Table 2

Ordered Logit Model - Question 5

	All responses and age/country controls	Three responses and age/country controls	Three responses and all controls
Total Number of Languages in Respondent's Country	-0.0058 *	-0.0064 *	0.0014
Age	-0.0011 *	-0.001 *	-0.0081 *
Country	0.0042 *	0.005 *	0.006 *
Total Number of Languages Squared			-0.001 *
Male (dummy)			0.0093
Urban (dummy)			0.3667 *

Note: * indicates p-value less than 1 percent

Figure 1 shows the graphical representation of model 1. There is very little change for the respondents who chose same, much better, or much worse. There is clear change in the respondents who chose better or worse.
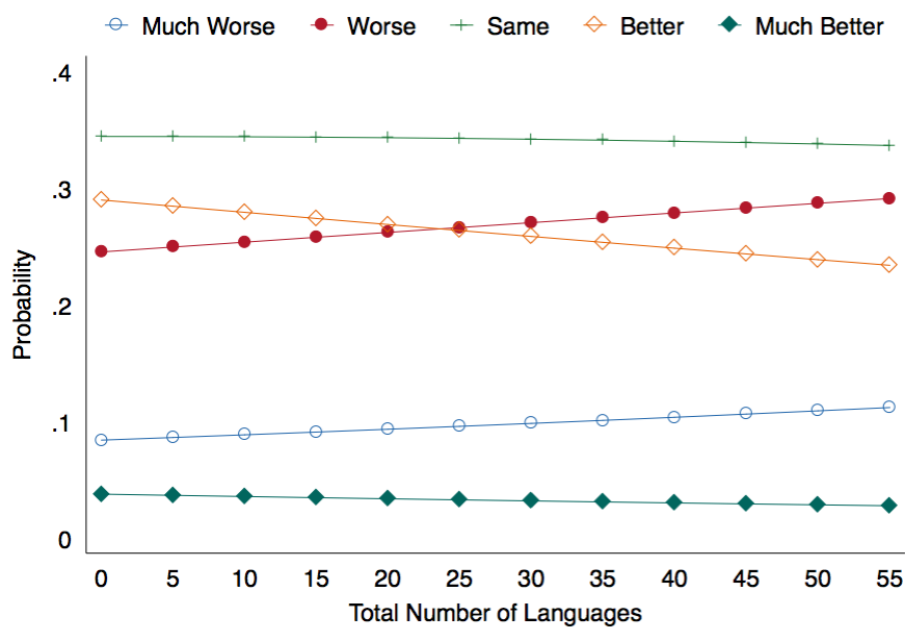
Figure 1

Figure 2 shows the graphical representation of model 2. The responses from the question have been recoded to create simple responses. Similar to model 1, there is very little change in respondents who selected "same", but larger differences for better or worse.
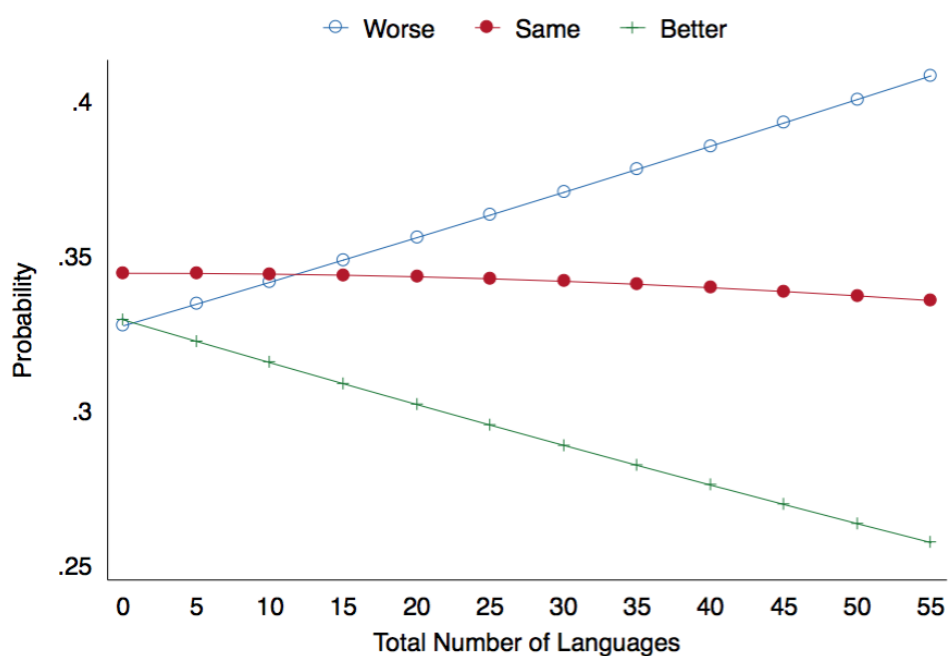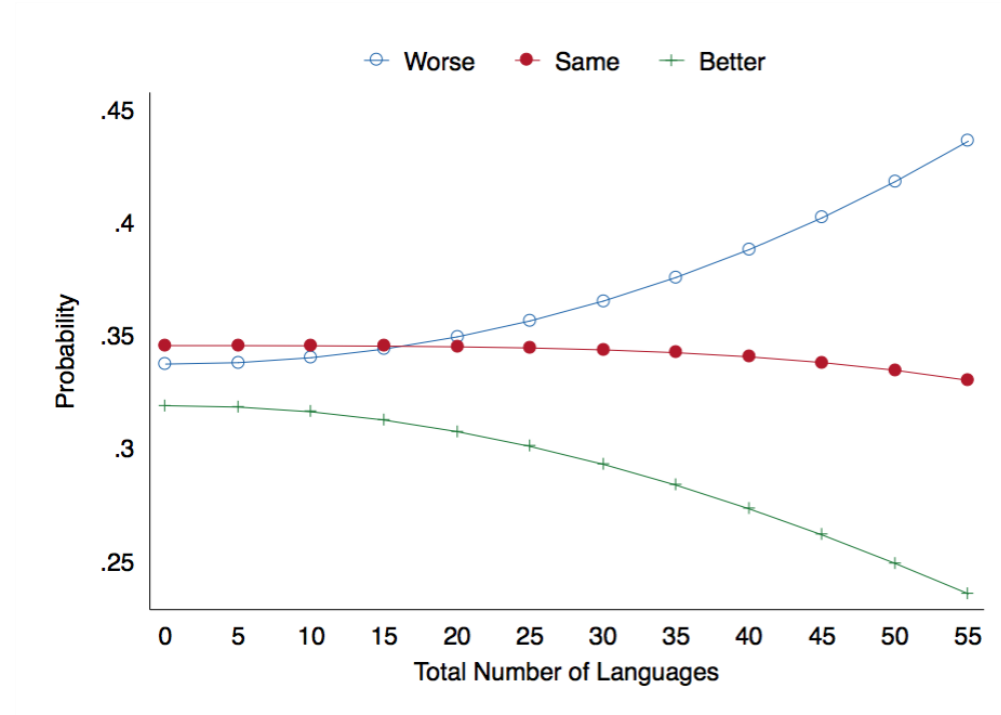
Figure 2

Figure 3 is the graphical representation of model 3. It uses the same recoded responses from model 2 while adding an interactive term for the continuous independent variable and controls for age, country, male, and urban location.

Figure 3

Limitations: There are few limitations to this analysis. It's possible that the data that was dropped (respondents who responded "Don't Know" or refused to respond) are different from the rest of the respondents in a meaningful way. If this is true, then 2.5% of the sample is missing and could alter the outcomes.

Another limitation is the graphical nature of language. Just because a country contains multiple homes that speak different languages doesn't necessarily mean that those homes are near each other and come into contact with one another. It's possible that they don't interact and therefore there can be no meaningful association between number of languages and views of others.

It's also not possible to prove causality with this analysis. There are most likely external factors that play a role in how someone views someone else's standard of living, which are not accounted for in this analysis.

Conclusions: This analysis would be much stronger with more control variables. Education is included in the survey, but it is categorical, which would be much better as a continuous variable. It would also be beneficial to have neighborhood level data so that comparisons can be made between people who actually interact with each other. The current analysis treats people who live in different parts of the same country as equal. Having neighborhood level data would also allow for more controls such as median income.

Despite the limitations, it does appear that there is an association between the number of languages and one's views of others. There is, however, no clear answer to the mechanism of the connection. Further research should explore what the driving cause of this connection is. Namely, are the countries that have more languages fundamentally different than the ones with fewer? Or, does speaking a different language inhibit one's ability to accurately assess the standard of living of someone else?