"I'm Not the President of Black America": Rhetorical Versus Policy Representation

Pavielle E. Haines, Tali Mendelberg, and Bennett Butler

Codebook and User Guide

2018

Table of Contents

General Introduction

The datasets described in this document can be used to evaluate Obama's rhetoric and policy representation of African Americans compared to past presidents. Data were taken from a variety of government sources, including the Office of Management and Budget, the Department of Labor, the U.S. Census Bureau, and the Bureau of Economic Analysis. Seven datasets are described in this document, two of which relate to presidential budget proposals, four of which relate to presidential rhetoric, and one of which relates to presidential DW-Nominate scores:

Yearly Presidential Budget Proposals: This dataset is contained in the file "obamayearlyspending.csv." Each row entry corresponds to presidential poverty and civil rights budget proposals for every year from 1970 through 2017.

Per Administration Presidential Budget Proposals: This dataset is contained in the file "obamaadminspending.csv." Each row entry corresponds to the average presidential poverty and civil rights budget proposals from a given administration from 1970 through 2017.

Yearly Presidential Rhetoric: This dataset is contained in the file "obamayearlyrhetoric.csv." Each row entry corresponds to measures of rhetorical content to the State of the Union address for every year from 1965 through 2016.

Per Term Presidential Rhetoric: This dataset is contained in the file "obamatermrhetoric.csv." Each row entry corresponds to the average rhetorical content of the State of the Union address given by a president during one term from 1965 though 2016.

Per Administration Presidential Rhetoric: This dataset is contained in the file "obamaadminrhetoric.csv." Each row entry corresponds to the average rhetorical content of the State of the Union address given by a president during his time in office from 1965 though 2016.

Intercoder Reliability: This dataset is contained in the file "obamaicreliability.csv." It includes coding decisions made by two different researchers who analyzed the content State of the Union addresses using the same set of instructions. This dataset can be used to evaluate the reliability of the coding scheme for measuring presidential rhetoric.

DW-Nominate Scores: This dataset is contained in the file "obamarollcallobject.rds." It is a matrix that includes data on every roll-call vote by every legislator through the 113th Congress.

Variables for Yearly Presidential Budget Proposals

(File Name: "obamayearlyspending.csv")

budgetid: The budget ID number. Each budget proposal is consecutively numbered, starting with the earliest budget proposal in the dataset, from one through forty-eight.

budgetyear: The year the proposed budget would take effect. The first budget year in the dataset is 1970 and the last is 2017. Budget proposals are offered nearly one year advance of when they would take effect. That is, the proposed budget for 2017 was offered to Congress in 2016.

propyear: The year the proposed budget was offered to Congress. The first proposal year in the dataset is 1969 and the last is 2016. Budget proposals are offered nearly one year advance of when they would take effect. That is, the proposed budget for 2017 was offered to Congress in 2016.

president: The name of the president who issued the budget proposal. Presidents in the dataset include Nixon, Ford, Carter, Reagan, H.W. Bush, Clinton, G.W. Bush, and Obama.

presidentid: The presidential ID number. Each president is consecutively numbered, starting with the earliest president in the dataset, from one to eight.

term: The term a president is in at the time they offer a budget proposal to Congress. This takes a value of either one or two.

obamadummy: A dummy variable indicating whether Obama was the president who issued the budget proposal. This takes a value of one for row entries associated with Obama and zero for every other president.

poverty: Total proposed poverty spending in millions of 2016 dollars. These amounts were calculated by taking the sum of proposed spending on anti-poverty programs that offered a disproportionate benefit to those in the first or second income quintiles. They reflect proposed spending totals for existing discretionary programs and proposed spending changes for mandatory programs, discretionary programs, and tax expenditures. For a more complete explanation of program selection and data collection, review the methods section in the paper and online appendices D, F, and G, or refer to the "Obama Data Sources and Collection Procedures.pdf" document included as part of the replication files.

civil: Total proposed civil rights spending in millions of 2016 dollars. These amounts were calculated by taking the sum of proposed spending on civil rights programs that offered a disproportionate benefit racial minorities. They reflect proposed spending totals for existing

discretionary programs and proposed spending changes for mandatory programs, discretionary programs, and tax expenditures. However, in practice, there no mandatory or tax-based civil rights programs. For a more complete explanation of program selection and data collection, review the methods section in the paper and online appendices H, or refer to the "Obama Data Sources and Collection Procedures.pdf" document included as part of the replication files.

povertydom: The percentage of the proposed discretionary domestic budget devoted to proposed discretionary poverty spending. This value is calculated by dividing the "povertydisc" variable (described below) by the "domestic" variable (described below).

civildom: The percentage of proposed discretionary domestic budget devoted to proposed discretionary civil rights spending. This value is calculated by dividing the "civil" variable (described above) by the "domestic" variable (described below).

perpoor: Proposed poverty spending per poor person. This does not reflect the actual benefit received by each poor person. This value is calculated by dividing the "poverty" variable (described above) by the "poor" variable (described below).

perunemployed: Proposed poverty spending per unemployed person. This does not reflect the actual benefit received by each unemployed person. This value is calculated by dividing the "poverty" variable (described above) by the "unemployed" variable (described below).

perblack: Proposed civil rights spending per African American. This does not reflect the actual benefit received by each African American. This value is calculated by dividing the "civil" variable (described above) by the "black" variable (described below).

gdp: Real GDP in billions of 2016 dollars in the proposal year.

gdpgrowth: Percent growth in real GDP in the proposal year.

deficit: The federal deficit in billions of 2016 dollars in the proposal year.

percentpoor: The percentage of Americans living below the poverty line in the proposal year.

percentunemployed: The percentage of unemployed Americans in the proposal year.

presdem: A dummy variable indicating whether the president was a Democrat. It takes a value of one for Democrats and zero for Republicans.

congressdem: A dummy variable indicating whether the Democrats controlled both the House and Senate in the proposal year. It takes a value of one when the Democrats control both chambers and zero otherwise.

electionbudget: A dummy variable indicating whether the president was up for reelection in the budget year. It takes a value of one for reelection years and zero for non-reelection years.

domestic: Total proposed discretionary domestic spending. This includes the president's proposed spending on every non-military, discretionary program.

povertydisc: Total proposed discretionary poverty spending. This variable is distinct from the "poverty" variable in that it excludes proposed changes to mandatory programs and tax expenditures.

poor: The number of Americans (in millions) living below the poverty line in the proposal year.

unemployed: The number of unemployed Americans (in millions) in the proposal year.

black: The number of African Americans (in millions) in the proposal year.

**Variables for Per Administration Presidential Budget Proposals
(File Name: "obamaadminpending.csv")**

president: The presidential administration that issued the budget proposals. Presidential administrations in the dataset include Nixon, Ford, Carter, Reagan, H.W. Bush, Clinton, G.W. Bush, and Obama.

presidentid: The ID number for the presidential administration. Each presidential administration is consecutively numbered, starting with the earliest president in the dataset, from one to eight.

poverty: The average total proposed poverty spending for all budget proposals offered by a presidential administration, in millions of 2016 dollars. This can be calculated by taking the average value of the "poverty" variable for each president in the Yearly Presidential Budget Proposals dataset.

civil: The average total proposed civil rights spending for all budget proposals offered by a presidential administration, in millions of 2016 dollars. This can be calculated by taking the average value of the "civil" variable for each president in the Yearly Presidential Budget Proposals dataset.

povertydom: The average percentage of the proposed discretionary domestic budget devoted to proposed discretionary poverty spending for all budget proposals offered by a presidential administration. This can be calculated by taking the average value of the "povertydom" variable for each president in the Yearly Presidential Budget Proposals dataset.

civildom: The average percentage of the proposed discretionary domestic budget devoted to proposed discretionary civil rights spending for all budget proposals offered by a presidential administration. This can be calculated by taking the average value of the "civildom" variable for each president in the Yearly Presidential Budget Proposals dataset.

perpoor: Average proposed poverty spending per poor person for all budget proposals offered by a presidential administration. This can be calculated by taking the average value of the "perpoor" variable for each president in the Yearly Presidential Budget Proposals dataset.

perunemployed: Average proposed poverty spending per unemployed person for all budget proposals offered by a presidential administration. This can be calculated by taking the average value of the "perunemployed" variable for each president in the Yearly Presidential Budget Proposals dataset.

perblack: Average proposed civil rights spending per African American for all budget proposals offered by a presidential administration. This can be calculated by taking the average value of the "perblack" variable for each presidential term in the Yearly Presidential Budget Proposals dataset.

**Variables for Yearly Presidential Rhetoric
(File Name: "obamayearlyrhetoric.csv")**

speechid: The ID number for each State of the Union speech. State of the Union speeches are consecutively numbered, starting with the earliest speech in the dataset, from one through forty-nine.

speechyear: The year the speech was given. The first speech year in the dataset is 1965 and the last is 2016.

refyear: The year immediately preceding each speech. State of the Union Speeches are generally given in January and are intended to review the country's progress in the preceding year.

president: The name of the president who gave the speech. Presidents in the dataset include Johnson, Nixon, Ford, Carter, Reagan, H.W. Bush, Clinton, G.W. Bush, and Obama.

presidentid: The ID number for each president. Presidents are consecutively numbered, starting with the earliest president in the dataset, from one to nine.

term: The term a president was in at the time of their speech. This takes a value of either one or two.

obamadummy: A dummy variable indicating whether Obama was the president who issued the speech. This takes a value of one for row entries associated with Obama and zero for every other president.

povertyper: The number of words devoted to the topic of poverty per every 10,000 words. This value is calculated by dividing the "povertyword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

ppovertyper: The number of words devoted to positive or neutral references to poverty per every 10,000 words. This value is calculated by dividing the "ppovertyword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

npovertyper: The number of words devoted to negative references to poverty per every 10,000 words. This value is calculated by dividing the "npovertyword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

civilper: The number of words devoted to the topic of race and/or civil rights per every 10,000 words. This value is calculated by dividing the "civilword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

pcivilper: The number of words devoted to positive or neutral references to race and/or civil rights per every 10,000 words. This value is calculated by dividing the "pcivilword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

ncivilper: The number of words devoted to negative references to race and/or civil rights per every 10,000 words. This value is calculated by dividing the "ncivilword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

middleper: The number of words devoted to the topic of the middle class per every 10,000 words. This value is calculated by dividing the "middleword" variable (described below) by the "words" variable (described below) and multiplying the quotient by 10,000.

gdp: Real GDP in billions of 2016 dollars in the year prior to the speech.

gdpgrowth: Percent growth in real GDP in the year prior to the speech.

deficit: The federal deficit in billions of 2016 dollars in the year prior to the speech.

percentpoor: The percentage of Americans living below the poverty line in the year prior to the speech.

percentunemployed: The percentage of unemployed Americans in the year prior to the speech.

presdem: A dummy variable indicating whether the president who gave the speech was a Democrat. Democrats take a value of one and Republicans take a value of zero.

congressdem: A dummy variable indicating whether the Democrats controlled both the House and Senate in the year the speech was given. Years where the Democrats controlled both chambers take a value of one and zero otherwise.

electionspeech: A dummy variable indicating whether the president was up for reelection at the end of the year the speech was given. Reelection years take a value of one and non-reelection years take a value of zero.

words: The total number of words in the speech

povertyword: The total number of words referencing poverty in the speech.

ppovertyword: The total number of words referencing poverty with a positive or neutral subtext.

npovertyword: The total number of words referencing poverty with a negative subtext.

civilword: The total number of words referencing race and/or civil rights in the speech.

pcivilword: The total number of words referencing race and/or civil rights with either a positive or neutral subtext.

ncivilword: The total number of words referencing race and/or civil rights in the speech with a negative subtext.

middleword: The total number of words referencing the middle class in the speech.

Variables for Per Term Presidential Rhetoric
(File Name: "obamatermrhetoric.csv")

president: The president giving State of the Union addresses during a presidential term. Presidents in the dataset include Johnson, Nixon, Ford, Carter, Reagan, H.W. Bush, Clinton, G.W. Bush, and Obama.

presidentid: The ID number for each president. Presidents are consecutively numbered, starting with the earliest administration in the dataset, from one to nine.

term: The term a presidential administration was in. This takes a value of either one or two.

obamadummy: A dummy variable indicating whether the State of the Union addresses in a particular term were given by Obama. This takes a value of one for row entries associated with the Obama and zero otherwise.

povertyper: The average number of words devoted to the topic of poverty per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "povertyper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

ppovertyper: The average number of words devoted to positive or neutral references to poverty per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "ppovertyper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

npovertyper: The average number of words devoted to negative references to poverty per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "npovertyper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

civilper: The average number of words devoted to the topic of race and/or civil rights per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average value of the "civilper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

pcivilper: The average number of words devoted to positive or neutral references to race and/or civil rights per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "pcivilper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

ncivilper: The average number of words devoted to negative references to race and/or civil rights per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "ncivilper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

middleper: The number of words devoted to the topic of the middle class per every 10,000 words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "middleper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

gdp: The average real GDP in billions of 2016 dollars for each presidential term. This value can be calculated by taking the average of the "gdp" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

gdpgrowth: The average percent growth in real GDP for each presidential term. This value can be calculated by taking the average of the "gdpgrowth" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

deficit: The average federal deficit in billions of 2016 dollars for each presidential term. This value can be calculated by taking the average of the "deficit" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

percentpoor: The average percentage of Americans living below the poverty line for each presidential term. This value can be calculated by taking the average of the "percentpoor" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

percentunemployed: The average percentage of unemployed Americans for each presidential term. This value can be calculated by taking the average of the "percentunemployed" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

presdem: A dummy variable indicating whether the presidential administration is Democratic in each term. This takes a value of one for Democratic administrations and zero otherwise.

congressdem: A dummy variable indicating whether the Democrats controlled both the House and Senate for at least half the years in the presidential term. This takes a value of one for Democratic control and zero otherwise.

electionspeech: A dummy variable indicating whether the president was up for reelection at the end of a term. This takes a value of one for reelection terms and zero for non-reelection terms.

words: The average total number of words for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "words" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

povertyword: The average total number of words referencing poverty for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "povertyword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

ppovertyword: The average total number of words referencing poverty with a positive or neutral subtext for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "ppovertyword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

npovertyword: The average total number of words referencing poverty with a negative subtext for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "npovertyword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

civilword: The average total number of words referencing race and/or civil rights for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "civilword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

ppovertyword: The average total number of words referencing race and/or civil rights with a positive or neutral subtext for State of the Union addresses given by a president during one term. This value can be calculated by taking the average value of the "pcivilword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

npovertyword: The average total number of words referencing race and/or civil rights with a negative subtext for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "ncivilword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

middleword: The average total number of words referencing the middle class for State of the Union addresses given by a president during one term. This value can be calculated by taking the average of the "middleword" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

Variables for Per Administration Presidential Rhetoric
(File Name: "obamaadminrhetoric.csv")

president: The president giving State of the Union addresses during a particular administration. Presidents in the dataset include Johnson, Nixon, Ford, Carter, Reagan, H.W. Bush, Clinton, G.W. Bush, and Obama.

presidentid: The ID number for presidential administrations. Each presidential administration is consecutively numbered, starting with the earliest administration in the dataset, from one to nine.

povertyper: The average number of words devoted to the topic of poverty per every 10,000 words for State of the Union addresses a president gave during their time in office. This value can be calculated by taking the average of the "povertyper" variable for each presidential term in the Yearly Presidential Rhetoric dataset.

ppovertyper: The average number of words devoted to positive or neutral references to poverty per every 10,000 words for State of the Union addresses given by a president during their time in office. This value can be calculated by taking the average of the "ppovertyper" variable for each presidential administration in the Yearly Presidential Rhetoric dataset.

npovertyper: The average number of words devoted to negative references to poverty per every 10,000 words for State of the Union addresses given by a president during their time in office. This value can be calculated by taking the average of the "npovertyper" variable for each presidential administration in the Yearly Presidential Rhetoric dataset.

civilper: The average number of words devoted to the topic of race and/or civil rights per every 10,000 words for State of the Union addresses given by a president during their time in office. This value can be calculated by taking the average of the "civilper" variable for each presidential administration in the Yearly Presidential Rhetoric dataset.

pcivilper: The average number of words devoted to positive or neutral references to race and/or civil rights per every 10,000 words for State of the Union addresses given by a president during their time in office. This value can be calculated by taking the average of the "pcivilper" variable for each presidential administration in the Yearly Presidential Rhetoric dataset.

ncivilper: The average number of words devoted to negative references to race and/or civil rights per every 10,000 words for State of the Union addresses given by a president during their time in office. This value can be calculated by taking the average of the "ncivilper" variable for each presidential administration in the Yearly Presidential Rhetoric dataset.

middleper: The number of words devoted to the topic of the middle class per every 10,000 words for State of the Union addresses given by a president during their time in office. This value can be calculated by taking the average of the "middleper" variable for each presidential administration in the Yearly Presidential Rhetoric dataset.

Variables for Intercoder Reliability
(File Name: "obamaicreliability.csv")

year: The year that the State of the Union speech being analyzed was given.

rhetoric: The type of rhetoric being analyzed. This takes five possible values: civil (total race and/or civil rights mentions); poverty (total poverty references); middle (total middle class references); negcivil (total negative references to race and/or civil rights); or negpoverty (total negative references to poverty).

coder1: The number of references associated with each category indicated by the "rhetoric" variable, as identified by the first researcher.

coder2: The number of references associated with each category indicated by the "rhetoric" variable, as identified by the second researcher.

Variables for DW-Nominate Scores
(File Name: "obamarollcallobject.rds")

n: The number of legislators.

m: The number of votes.

legis.data: A data frame containing information for each legislator, including their ICPSR code, name, party, state, and official DW-Nominate scores.

vote.data: A data frame containing information for each roll-call vote, including its identification number, short description, clerk rollcall number, the vote question, and the amendment author (if available).