

# 1 Classify a Model

I chose to read an article from the Journal of Political Science that studied the affect of the Voting Rights Act (VRA) Language Assistant Provision (LAP) on the representation of Latinos in the United States political system, specifically in local school boards. Briefly, the LAP in the VRA stipulates that municipalities must provide registration, voting notices, forms, instructions, assistance, ballots, or other materials relating to the electoral process in the language of minority group as well as in English. There has been evidence of increased voter participation as a result of these provisions but no studies have looked at if there is also increased representation of minority populations as well. This is the citation for the article:

Marschall, M. J. and Rutherford, A. (2016), Voting Rights for Whom? Examining the Effects of the Voting Rights Act on Latino Political Incorporation. American Journal of Political Science, 60: 590606. doi: 10.1111/ajps.12182

The authors created a model that was broken down into two phases where the first phase modeled the first Latino being elected in a district (they call this the "hurdle") and the second phase modeled one more more Latinos being elected (they call the "expansion") in a district. Thus the model looks like this:

$$g(y|\theta) = \begin{cases} f_1(0) & y = 0 \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(y|\theta) & y \geq 1 \end{cases}$$

The authors model these two stages via a logit model. They model the second stage via a zero-truncated Poisson model over the subset of observations with 1 or more Latinos represented in the School Board . Since there are two stages of representation there are two endogenous variables in the model (1) the incidence of representation and (2) the extent of representation. They use many exogenous variables, the primary one being if the district is covered under VRA. Others include: percentage of voting-age hispanics, majority white district (indicator variable), school board size, percentage of seats elected at-large, percentage of Latinos who live in owner-occupied housing, percent of latinos who earned at least a bachelor's degree, percent foreign born, percent speaking spanish only, percent of voting age blacks, percent of white college educated people. Additionally a set of time dummy variables were included to investigate whether Latino board representation has changed over time. These variables are spaced at four-year intervals from 1984 to 2012. Finally some control variables were included like total population size, median home value, and percent of people in poverty.

The model presented in this paper is a dynamic, linear, deterministic model. The authors use a Poisson and Logit model both of which are generalized linear models. It is dynamic because the authors include a set of dummy variables of four-year intervals from 1984 to 2012 to investigate whether Latino political representation has changed over time which allows the model to account for time-dependent changes. It is deterministic because variables are recorded from observations rather than calculated through probability distributions.

In addition to variables included in this model one variable I think the authors should include is a measurment of how well the LAP of the VRA is enforced. They could do this by requesting sample ballots in minority languages in each district and seeing if one is returned. While the VRA is a law (though parts of it have been overturned by the Supreme Court in 2013) local municipalities might not actually implement the law correctly. It is noted in the article that it is a hard provision for the government to enforce.

Note: Findings from this study indicate that the language assistance provisions of the Voting Rights Act had meaningful effect on Latino school board representation.

## 2 Create a Model

A basic model for predicting the lifespan of popular musician is below:

$$lifespan = \beta_1 1960s + \beta_2 druguser + \beta_3 num\_records\_sold + \beta_4 avg\_life\_expectancy + \epsilon$$

There are several key factors in determining the lifespan of a musician, however in general the life expectancy of a musician probably is probably similar to the life expectancy of any other citizen with some added variables that might affect the life expectancy of a musician specifically. Thus I wanted to include an avg life expectancy variable that would be a lookup on an actuarial table such as this one by the world bank link. This table gives the average life expectancy in years of someone born on a specific year in a specific country and should control for the conditions the musician lives in such as the healthcare system of their home country.

Next, I tried to include variables that would affect the health and lifespan of musicians specifically. I included an indicator variable for 1960s which will be 1 if the musician was prominent in the 1960s and 0 otherwise. The reason I included this variable is because during the 1960s drug use was very prevalent among musicians including many hard drugs such as LSD and which are very dangerous. I also included an indicator variable if the musician was a known drug user or smoker. While this can be hard to know for sure, one might be able to tell by looking at news articles or commentary on the musician by other celebrities. This would affect the health and life expectancy of a musician also.

I also included some musician specific variables such as how many records they have sold. I'm not sure if this will have any affect but I would like to try it to see the result. I would think the musician would have to have a certain amount of fame for any of these musician specific variables to come into play. Thus after experimentation, it might make sense to have a two-phase model like the one I studied in this paper. It would work like this: the life expectancy of a musician is the same as a non-musician if they have sold under a certain amount of records while it would have these "musician effects" if they sold more than a certain amount of records. I also think their lifespan could depend on the number of records they made (fame could be a proxy for other variables such as how much travel they have done, or how many years they were successful).

There are a lot of variables I chose not to include such as the health care quality of the country and the year that the musician was born. I did this because I think that the worldbank table will and research has shown that several key variables in predicting the life expectancy are birth country, birth year, and education level so I included all three of those variables. In addition I included some specific variables I think might help predict the lifespan of a musician specifically I think that musicians might be more likely to die of a drug or alcohol overdose so I included that variable.

In order to do a preliminary test to determine if any of these factors were relevant, the hardest part would be acquiring data. I would start by trying to find data from an almanac or wikipedia to build out a dataset I could use. I would get a sample of a few thousand musicians that have passed away and look at their bios to code some of the other variables. I would also start looking for some more clues such as reading about the cause of death of these musicians to determine if there were any patterns. Coding some of the data would be difficult but it would be an easy task to add to aws mechanical turk. Another data sources to find the number of records sold would be from billboard.com or an almanac. After creating the dataset for a few thousand musicians I would run a model in R. If the results looked promising I would work on acquiring more data.